The Dissertation Committee for Liang Xiang Certifies that this is the approved version of the following dissertation:

EXPANDING THE GENETIC CODE IN MAMMALIAN CELLS

Committee:

Zhiwen Jonathan Zhang, Supervisor

George Georgiou, Co-Supervisor

Krishnendu Roy

Pengyu Ren

Whitney Yin

EXPANDING THE GENETIC CODE IN MAMMALIAN CELLS

by

Liang Xiang, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2011

Dedication

To my family; both close and afar. You’re always there to help shine a light. Your advice has always helped to set things right. Thank you for all your love.

Acknowledgements

I am very grateful to my supervisor Dr. Zhiwen Jonathan Zhang. You have been both a mentor and a friend. Thank you for sharing your advice and experiences. I also appreciate the help of Dr. Jie Zhu and Dr. Aiko Umeda who have taught me everything I know about molecular cloning. To Dr. Gabrielle Thibodeaux and Dr. Kathryn Moncivais thank you for training me in all things related to mammalian culturing. I am also thankful for the other members of the Zhang lab, all the undergraduates, graduate students, and post-doctorates that have come and gone but left their mark.

Thank you to the University of Texas at Austin for the educational experience. Thanks to all her faculty and staff, the professors who have shared their wisdom and passion, the technicians and trained personnel who are always knowledgeable and eager to assist. Thank you to Mr. Bobby Knight and Dr. Pratibha Saxena for demonstrating the teaching process. For providing me a stage to practice, to fail, and to succeed. Thank you to my committee members and the Biomedical Engineering Department for their continuous guidance and direction.

Finally, thanks to my friends for their constant support. To the group of five from Katy, Texas. We bonded together while we were young. Grew up and matured while living apart. Thank you for these past fifteen years, for always being the ones I turn to first.

Expanding the Genetic Code in Mammalian Cells

Liang Xiang, Ph.D.

The University of Texas at Austin, 2011

Supervisor: Zhiwen Jonathan Zhang, George Georgiou

Proteins are diverse polymers of covalently linked amino acids. They play a role in almost every biological process that occurs within an organism. Twenty different amino acids are genetically encoded by mammalian cells to build proteins. The sequence of these amino acids determines the protein's final shape, structure, and function. Modern molecular cloning techniques allow for the genetic encoding and expression of mutant proteins that have one or more amino acids replaced with one of the others. The roles of individual amino acids in a protein can therefore be studied. Proteins with novel functions have also been designed or evolved using this technology. However, the genetic code is limited to the twenty natural amino acids.

Nonnatural amino acids have unique side groups that not found on any of the twenty natural amino acids. They can be site-specifically incorporated using a mutant orthogonal suppressor tRNA/aminoacyl-tRNA synthetase (aaRS) pair. Each pair only allows for one type of nonnatural amino acid to be genetically encoded. This technology has resulted in the incorporation of over fifty different types of nonnatural amino acids into proteins in *prokaryotic* and *eukaryotic* cells. Unfortunately, most of these pairs are not orthogonal outside of *prokaryotic* systems and only a few have been developed for mammalian cells. To create more mammalian pairs a nonnatural aaRS has to be evolved and screened in a cumbersome process. In this dissertation an approach is outlined that

can be used to change the orthogonality of existing nonnatural suppressor tRNA/aaRS pairs. As a result of the orthogonality change many previously unavailable pairs can be shuttled into mammalian cells. The ability to genetically encode a 21st amino acid is a powerful tool in the study and engineering of proteins.

Table of Contents

List of Figures

Chapter One

Nonnatural Amino Acid Containing Proteins

1.1 INTRODUCTION

All living organisms use amino acids as the basic building blocks to assemble proteins. These proteins are defined by a unique sequence of covalently linked amino acid residues which determines its final structure and function. Discovery of the genetic code has revealed the mechanism by which DNA is used to store a protein's specific sequence. Twenty natural amino acids that contain different side chains can be genetically encoded this way. Advancements in molecular cloning techniques have allowed us to express any amino acid sequence by encoding the information within a DNA template. In the past few decades, progress has also been made in the design and expression of proteins containing nonnatural amino acids [1]. These are usually analogs of their natural counterparts but contain atypical side chains (**Figure 1.1**) with unique chemical, fluorescent, or physical attributes. Nonnatural amino acid can be used to dramatically affect the properties and function of a protein.

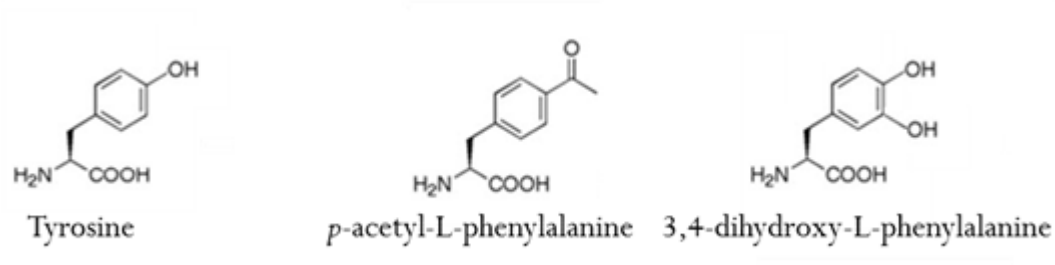


Figure 1.1: Structure of tyrosine and the nonnatural amino acids *p*-acetyl-L-phenylalanine and 3,4-dihydroxy-L-phenylalanine (L-DOPA). Although they are have a similar backbone, nonnatural amino acids have atypical side chains that contain functional groups which are not found on any of the twenty natural amino acids.

Nonnatural amino acids can be a powerful tool in advancing our understanding of proteins. Evidence suggests that cells require more diverse building blocks than the natural amino acids. For example, post-translational modification processes are often employed by cells to chemically modify one or more side chains on a protein [2]. Nonnatural amino acids can be used to mimic these modifications in order to study their purpose and effects. Protein engineering is another field that can greatly benefit from the use of nonnatural amino acids [3]. Roger Tsien's group have engineered a number of GFP variants with novel fluorescent properties by substituting various residues in the GFP with one of the other twenty [4]. Side chains can be directly involved in the functionality of the protein and there exists a lot more diversity between the nonnatural amino acids. Therefore, they can allow more possible protein designs. This chapter will review some of the current methods of incorporating nonnatural amino acids into a protein and some of their applications.

1.2 Protein Studies Using Nonnatural Amino Acids

Nonnatural amino acids can be incorporated into a protein using chemical or biosynthetic methods. *In vitro* or *in vivo* biosynthetic techniques are usually preferred because chemical synthesis is costly and cannot be used to produce large proteins [5]. These methods involve the manipulation of the aminoacylation reaction when a tRNA is covalently attached (charged) to an amino acid. A nonnatural amino acid charged tRNA can be used to incorporate a nonnatural amino acid during protein translation. After mRNA is transcribed from a DNA sequence, the transcript is translated into an amino acid sequence by the ribosome complex. Correct base pairing between the mRNA (codon) and charged tRNA (anticodon) regions will result in the transfer of the attached amino acid to the growing peptide chain (**Figure 1.2**). The ribosome does not proofread the attached amino acid so a mischarged tRNA can result in the wrong amino acid being added to a protein. Two different biosynthetic approaches take advantage of this phenomenon to incorporate nonnatural amino acids into proteins.

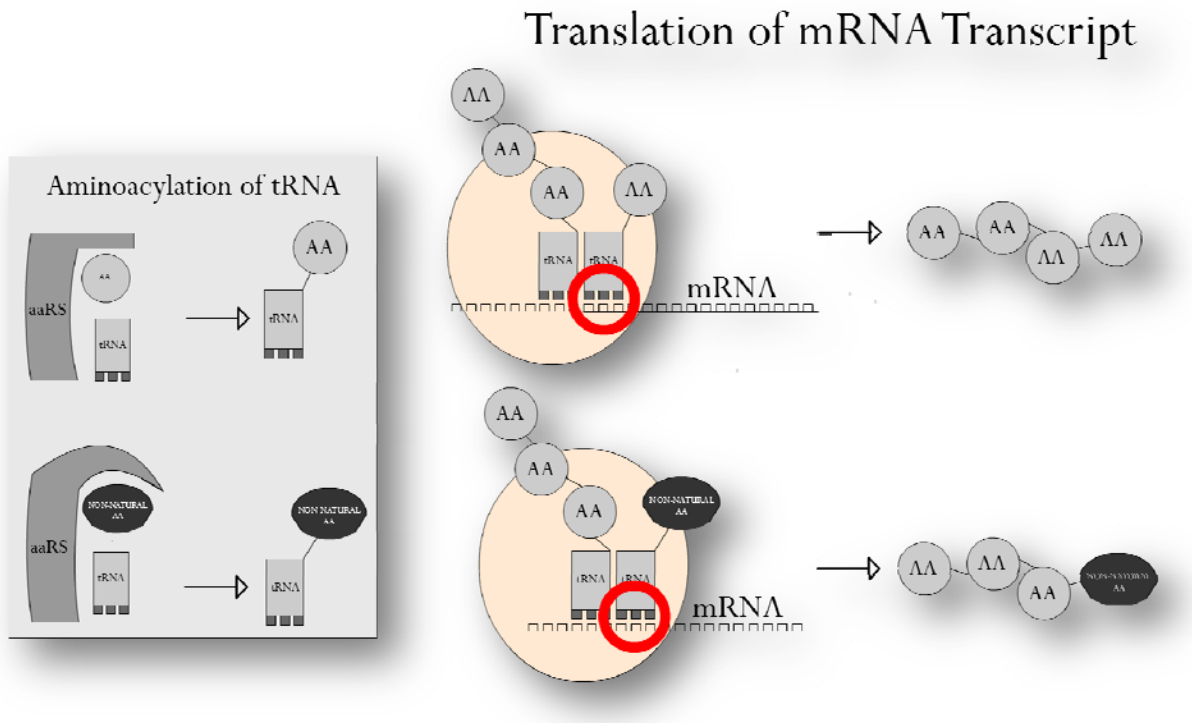


Figure 1.2: Aminoacyl-tRNA synthetases (aaRS) catalyze the covalent attachment of an amino acid to a corresponding tRNA. Biosynthetic methods techniques use tRNAs that have been charged with a nonnatural amino acid. Correct base pairing between the mRNA and tRNA (red circle) is the only requirement for the attached amino acid to be added to the growing peptide chain. (Adobe Flash CS5)

Residue-specific incorporation uses endogenous cellular machinery to replace one of the twenty natural amino acids in the genetic code with a nonnatural amino acid [6]. This has proven useful when one wants to incorporate a nonnatural amino acid into all proteins within the host. This can also greatly affect the overall properties of the protein since several residues are usually replaced (**Figure 1.3**). Site-specific incorporation expands the genetic code through the use of mutant orthogonal suppressor tRNA/aminoacyl-tRNA synthetase (aaRS) pairs. Each pair corresponds to a specific nonnatural amino acid. Proteins composed of twenty one amino acids can therefore be

produced, and the location of the nonnatural amino acid genetically encoded. The effect of the nonnatural amino acid on the protein can be minimized since it is genetically encoded (**Figure 1.3**). This can allow a nonnatural amino acid incorporated protein to be expressed without affecting other cellular proteins [3]. In the following sections we review the science and methodology behind both biosynthetic approaches as well as highlighting some of the applications currently demonstrated by using these techniques.

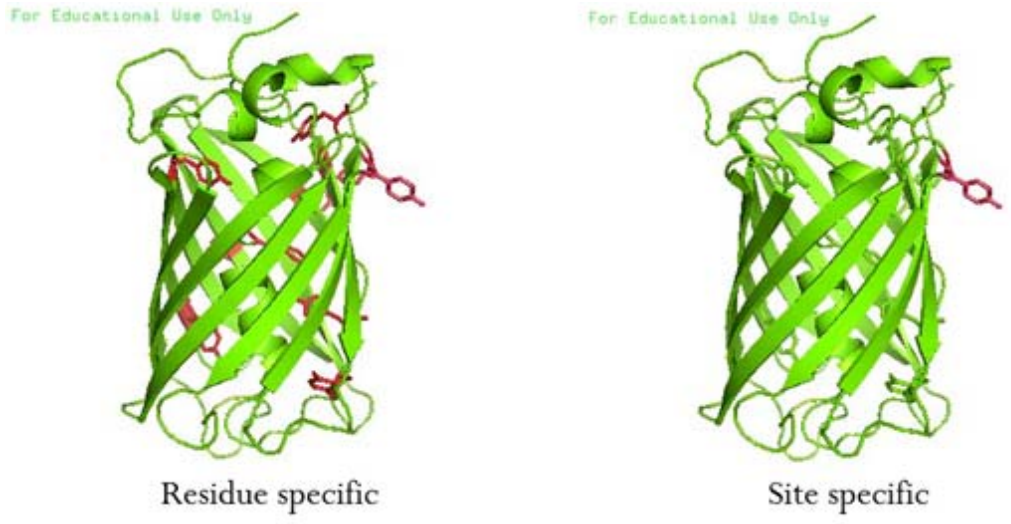


Figure 1.3: Residue specific incorporation replaces one type of amino acid (i.e. tyrosine shown in red) with the nonnatural amino acid. As a result, all residues within the protein are replaced resulting in dramatic changes in the overall properties of the protein. In addition, all cellular proteins will the host will be affected. Site-specific incorporation can be used to target the nonnatural amino acid to a specific location within the protein, this allows for more fine tuned modifications. Images were created with the PyMOL Molecular Graphics System (pymol.org)

1.2.1 Residue-specific Incorporation

Some natural aminoacyl-tRNA synthetases (aaRS) are promiscuous in identifying their target amino acid [7]. These wild type aaRS can be tricked into charging their cognate tRNAs with a nonnatural amino acid analog if it is supplied at a high enough concentration. Cohen and coworkers first demonstrated this in the 1950's when they quantitatively incorporated analogs such as seleno-methionine [8], nor-leucine, and para-fluorophenylalanine [9] in place of their natural amino acids. All amino acids of a chosen type in the cell will have an equal chance of being replaced by the nonnatural amino acid. There is no genetic manipulation required since endogenous tRNA/aaRS pairs are used. Furthermore, nonnatural amino acids can be incorporated into any protein within a cell without having to know its sequence [6].

Numerous studies have shown the various applications of residue-specific incorporation. Dieterich *et al.* devised a bio-orthogonal noncanonical (nonnatural) amino acid tagging method (BONCAT) to pulse label newly synthesized proteins with the nonnatural amino acid azidohomoalanine [10]. Cells are first grown in normal medium which results in the synthesis of proteins containing the twenty natural amino acids. When alanine was replaced with azidohomoalanine in the growth media, all newly translated proteins will have an azidohomoalanine in place of an alanine. The azide functional group is not present on any of the natural amino acids, thus these nonnatural amino acid containing proteins can be selectively purified and identified. Another temporal labeling application was demonstrated by Tirrell and coworkers when they used two methionine analogs (azidohomoalanine and homopropargylglycine) to pulse-label proteins at two different time points in mammalian cells [11]. As with the one nonnatural amino acid labeling technique, cells were first grown in normal medium and then switched to media where methionine was replaced with azidohomoalanine followed by

homopropargylglycine. Fluorescent molecules can be used to label these proteins by covalently linking them to the unique functional groups on the nonnatural amino acids. Proteins synthesized when the nonnatural amino acid was present in the medium would contain it in place of alanine. They demonstrate the use of nonnatural amino acids as an easier and safer alternative to conventional methods that often use radioactive materials [12].

Another area of application for residue-specific incorporation is the production of proteins with modified properties [6]. Since multiple resides within a protein are usually replaced with the nonnatural amino acid, the folding and functional characteristics of the protein can be greatly affected. The melting temperature of a collagen peptide was shown to increase by more than 50 ºC when nonnatural amino acids were substituted in place of proline [13]. Shoulders *et al.* replaced all of the proline residues within the collagen motif, proline-proline-glycine, with 4-fluoroproline and 4-methylproline. The resulting nonnatural amino acid containing collagen peptide still formed a triple helix structure that was similar to natural collagen but exhibited much better thermostability. Another example is when Yoo and coworkers replaced all the leucine residues in green fluorescent protein with 5,5,5-trifluoroleucine. Then by using directed evolution, they were able to isolate a mutant that contained fifteen residues of 5,5,5-trifluoroleucine. Even though it contained several nonnatural amino acids it still had similar folding and fluorescence properties to that of wild type GFP [14]. Lepthien *et al.* created a blue fluorescent human annexin protein when they replaced a tryptophan residue with 4-azatryptophan [15]. This demonstrated a very simple way to directly visualize and track the location of a protein without having to covalently attach a fluorescent molecule [16]. The ability of nonnatural amino acids to affect the characteristics of a protein's active site is also an area of research.

Residue-specific incorporation is limited by the fact that it relies on the low activity of endogenous amino-tRNA synthetases (aaRS) for natural amino acid analogs. The nonnatural amino acid must be supplied either in replacement or at a much higher concentration than the natural amino acid. Auxotrophic strains are commonly used to improve the incorporation efficiency by minimizing background [17].The fidelity of endogenous aaRS can also be reduced by mutating its active site to increase the activity of the synthetase for the nonnatural amino acid [18]. Although global incorporation can be beneficial in studies where the modification of all proteins is required, too much nonnatural amino acid content could very likely have a negative impact on protein folding and function. Proteins other than the one of interest would also have every amino acid of one type replaced by the nonnatural amino acid. BONCAT can be used as a tool to isolate newly synthesized proteins under various conditions, but the global modification of all cellular proteins limits the significance of such studies.

1.2.2 Site-specific Incorporation

Site-specific incorporation can be used to add a single nonnatural amino acid anywhere within a protein without globally affecting the host cell's proteins. A method was developed, independently, by Schultz [19] and Chamberlain [20] to genetically encode a nonnatural amino acid using a stop codon. All three nucleotide combinations are used by the cell to encode for one of the twenty natural amino acids. Only the three stop codons amber (TAG), ochre (TAA), and opal (TGA) do not encode an amino acid and instead are used to signal translation termination. However, if an aminoacylated tRNA with a mutation in the anticodon region can correctly base pair with the stop codon then protein translation will continue [21]. Any stop codon that appears in a gene could be

effectively ‘suppressed’ as long as the cell has a charged suppressor tRNA. So, they first modified a tRNA gene in its anticodon sequence (GUA to CUA) and then transcribed it *in vitro*. This ‘suppressor’ tRNA folded correctly and was chemically aminoacylated with a phenylalanine analog. Next, an mRNA construct was created for a gene of interest. A codon within the gene was changed to the amber stop codon (TAG). This would normally cause translation to halt, but the suppressor tRNA can base pair with the TAG sequence and allow translation to continue. The ribosome lacks the ability to proofread the identity of the amino acid on the tRNA and will transfer it as long as correct base pairing occurs between the tRNA anticodon and mRNA codon sequences. These components were used with an *in vitro* cell-free translation system in order to incorporate several phenylalanine analogs site-specifically [19]. However, the amount of protein generated by this approach is limited since a tRNA cannot be recharged once it has transferred the attached amino acid. Any protein with a known DNA sequence can be modified with a nonnatural amino acid using this method.

This technique was next expanded to *in vivo* experiments when an amber suppressor tRNA was aminoacylated with a tyrosine analog and injected into *Xenopus* oocytes [22]. An mRNA constructs containing the gene for the nicotinic acetylcholine receptor had a codon replaced with the amber stop codon and was also injected into the oocytes. As a result, they demonstrated *in vivo* site-specific incorporation of a nonnatural amino acid in response to a stop codon. But this technique was still limited by the amount of nonnatural amino acid containing proteins produced since there was no way to regenerate (recharge) the suppressor tRNA in the cells. Schultz later resolved the tRNA limitation and applied his approach to *in vivo* systems when he developed and introduced an orthogonal suppressor tRNA/aminoacyl-tRNA synthetase (aaRS) into cells as a 21st pair [23]. They identified a tyrosyl-tRNA/tyrosyl-tRNA synthetase (TyrRS) from

Methanococcus jannaschii that was orthogonal in *Escherichia coli* cells [24]. Orthogonality of this pair ensured that endogenous *E. coli* TyrRS would not aminoacylate the *M. jannaschii* tRNA nor would the *M. jannaschii* TyrRS aminoacylate any *E. coli* tyrosyl-tRNA. The gene for *M. jannaschii* tyrosyl-tRNA was modified in its anticodon sequence from GUA to CUA to allow it to base pair to the amber stop codon (TAG). The *M. jannaschii* TyrRS's active site was then mutated to change its specificity from tyrosine to a nonnatural amino acid [23]. They identified five residues within the enzyme's active site (Tyr32, Glu107, Asp158, Ile159, and Leu162) responsible for recognizing the amino acid side chain. A plasmid library of TyrRS mutants was created by randomly mutating all five locations. Cells transformed with the library were screened with a round of positive selection (**Figure 1.4**) to identify mutants that could recognize a tyrosine analog (*O*-methyl-L-tyrosine). Negative selection was then used to remove mutants that could still recognize tyrosine. Alternating rounds of positive and negative selection he most active and specific mutant TyrRS to be identified. To incorporate the nonnatural amino acid site-specifically, they mutated the third codon of a gene for dihydrofolate reductase to the amber stop codon. *E. coli* cells were transformed with a plasmid containing the mutated dihydrofolate reductase gene, a gene for the suppressor tRNA, and the nonnatural aaRS gene. When the cells were supplied with a nonnatural amino acid in the media, the *E. coli* cells were able to produce mutant dihydrofolate reductase proteins containing *O*-methyl-L-tyrosine in the third position [23].

Each nonnatural amino acid requires a new nonnatural aaRS. Several of these pairs have been evolved to incorporate many nonnatural amino acids. Two pairs of interest are the ones for the nonnatural amino acids *p*-acetyl-L-phenylalanine [25] and 3,4-dihydroxy-L-phenylalanine [26]. Both nonnatural aaRS were evolved from *M. jannaschii* TyrRS. *P*-acetyl-L-phenylalanine (**Figure 1.1**) contains a novel ketone

functional group not found on any of the other twenty natural amino acids. Using the method described above, Wang *et al.* evolved a mutant TyrRS with > 99.8% fidelity for *p*-acetyl-L-phenylalanine. The high activity of this mutant TyrRS was also demonstrated when they were able to express and purify 3.6 mg of a protein containing *p*-acetyl-L-phenylalanine from 1 liter of *E. coli* culture (compared with 9.2 mg/L with a when using wild type TyrRS) [25]. Furthermore, they labeled the nonnatural amino acid containing protein by covalently attaching a fluorescein hydrazide to the ketone group *in vitro*. They also hypothesized that proteins with unique chemical functions could be generated by using the ketone containing amino acid. A nonnatural aaRS derived from *M. jannaschii* TyrRS was evolved that had specificity for 3,4-dihydroxy-L-phenylalanine (L-DOPA) [26]. Using this mutant TyrRS, Alfonta *et al.* expressed 1 mg/L of a L-DOPA incorporated protein (compared to 5.6 mg/L with a wild type TyrRS). L-DOPA incorporated protein is beneficial because it can also be labeled *in vitro* by using the chemical properties of the nonnatural amino acid. It can also be used to study and capture transient protein-protein interactions. Umeda *et al.* was able to show the dimerization of a sortase A protein from *Staphylococcus aureus* by using L-DOPA to site-specifically crosslink the weak dimer complex [27].

In addition to *p*-acetyl-L-phenylalanine and 3,4-dihydroxy-L-phenylalanine, more than 30 other nonnatural amino acids have also been incorporated into proteins using this technique [1]. For example, a *p*-azidophenylalanine incorporated human superoxide dismutase-1 enzyme was site-specifically attached to a PEG group (PEGylation) [28]. This can be used to design new therapeutic proteins with improved solubility and increased serum half-life [29]. Bose *et al.* was able to photo-regulate the DNA binding affinity of a protein by incorporating a nonnatural amino acid [30]. An isoleucine in a catabolite activator protein (CAP) was replaced with the nonnatural amino acid

phenylalanine-4-azobenzene. Upon stimulation with UV-light, phenylalanine-4-azobenzene converted from a *trans* to a *cis* configuration. As a result, the mutant CAP's binding constant drop by almost four folds. Another example is the use of the nonnatural amino acid *p*-iodo-L-phenylalanine to help determine the crystal structure of a protein. A phenyalanine in bacteriophage T4 lysozyme was replaced with *p*-iodo-L-phenylalanine. The resulting protein was purified and its structure determined with much greater ease than the wild type protein which did not contain *p*-iodo-L-phenylalanine [31]. These examples demonstrate the vast number of applications for site-specifically incorporating nonnatural amino acids into proteins.

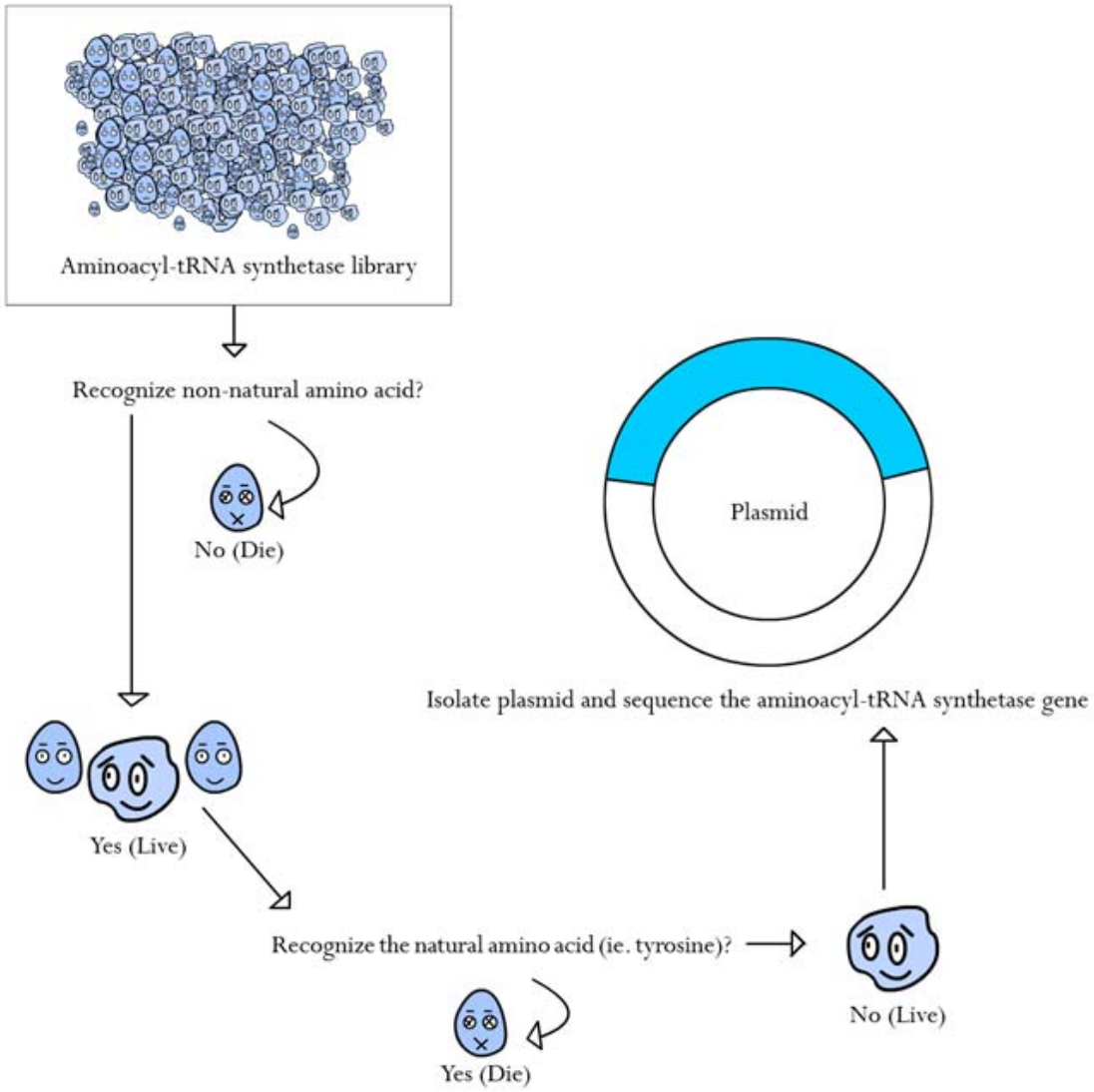


Figure 1.4: A genetic library is created containing random mutations in the aminoacyl-tRNA synthetases's active site. These plasmids are cloned into cells who are then subjected to several rounds of positive and negative screening in order to identify mutants capable of recognizing the nonnatural amino acid without recognizing the natural amino acid. The plasmid is isolated from surviving clones and then sequenced. Normally, upwards of 10^8 clones must be screened in order to identify a few mutants. (Adobe Flash CS5)

1.3 SITE-SPECIFIC INCORPORATION IN MAMMALIAN CELLS

Nowak was the first to apply the stop codon suppression method of nonnatural amino acid incorporation to a mammalian system. Suppressor tRNAs were chemically charged with nonnatural amino acids and injected into a *Xenopus oocyte*. An mRNA construct containing an amber codon was also injected into the cell. Several nonnatural amino acids were added into ion channels using this technique [22]. However, only very small amounts of nonnatural amino acids can be incorporated this way since the suppressor tRNAs cannot be regenerated (recharged). Sakamoto *et al*. overcame this limitation when he used a mutant *E. coli* TyrRS to enzymatically charge a *Bacillus stearothermophilus* tyrosyl-tRNA derived suppressor tRNA within a mammalian cell [32]. Introduction of this orthogonal pair in CHO-Y cells was used to incorporate 3-iodo-L-tyrosine in response to an amber stop codon. Two amino acids were changed in the *E. coli* TyrRS's active site to create a mutant that could recognize preferentially recognize 3-iodo-L-tyrosine over tyrosine (95% efficiency).

Schultz's group has also evolved several nonnatural aaRS. A survival assay (**Figure 1.4**) was used to screen 10^8 variations of *E. coli* TyrRS that contained random mutations in its active site [33]. They were able to evolve five different nonnatural aaRS that are specific (>99.8%) for *p*-acetyl-L-phenylalanine, *p*-benzoyl-L-phenylalanine, *p*-azido-L-phenylalanine, *O*-methyl-L-tyrosine, and *p*-iodo-L-phenylalanine. An *E. coli* tyrosyl-tRNA derived suppressor tRNA was paired with the nonnatural aaRS and introduced into yeast cells. As a result, a nonnatural amino acid could be incorporated in response to a stop codon. Wang *et al.* also demonstrated that these yeast evolved pairs could be used in mammalian systems. HeLa, mouse hippocampal neuron, and HEK293T cells were able to incorporate *O*-methyl-L-tyrosine in response to an amber stop codon when provided with the nonnatural aaRS and suppressor tRNA pair [34]. Finally, they showed that replacing a

tyrosine with *O*-methyl-L-tyrosine in a voltage-dependent K+ channel greatly decreased the inactivation of the channel. *P*-acetyl-L-phenylalanine and *p*-benzoyl-L-phenylalanine were also successfully incorporated into a protein in mammalian cells (HEK293T/HEK293S) by using the yeast evolved pairs [35]. These ketone containing nonnatural amino acids were incorporated into two G protein-coupled receptors and subsequently labeled with fluorescein hydrazide.

An orthogonal pair from *Bacillus subtilis* was used to genetically encode a nonnatural amino acid, 5-hydroxy-L-tryptophan, in HEK293T cells [36]. The bacterial pair from *B. substilis* is already orthogonal in mammalian cells, but Zhang *et al.* had to mutate the tryptophan-tRNA sequence in order to improve its *in vivo* transcription. One residue within the tryptophan-tRNA synthetase (TrpRS) was mutated resulting in a mutant TrpRS with >97% specificity for 5-hydroxy-L-tryptophan. The fluorescence properties of a bacteriophage T4 fibritin (foldon) was then altered when a tryptophan was substitute for 5-hydroxy-L-tryptophan.

Although many different types of nonnatural amino acids have been successful incorporated into a protein in *prokaryotic* systems, only a fraction of them have been incorporated in mammalian cells. The development of additional pairs will require substantial amounts of time and effort in order to either rationally design a nonnatural aaRS [36] to evolve one [37]. Since each pair can only be used to incorporate one nonnatural amino acid, in order to expand their application in mammalian cells more pairs will have to be developed.

1.4 Overview and Specific Aims

1.4.1 Overview

Site-specific incorporation of nonnatural amino acids is a powerful tool in the study and engineering of proteins. Orthogonal pairs consisting of a suppressor tRNA and nonnatural aaRS can be used to incorporate large amounts of nonnatural amino acids into proteins in response to the amber stop codon. Genetic encoding allows the site-specific incorporation of the nonnatural amino acid within a protein. This prevents the global incorporation that is seen when using residue-specific techniques. However, engineering a nonnatural aaRS through rationale design or directed evolution is a difficult process. Thus far, less than a dozen nonnatural aaRS have been evolved for use in mammalian cells. On the other hand, a plethora of *M. jannaschii* derived nonnatural aaRS [23, 26, 38] have been designed but cannot be used in an *eukaryotic* system due to non-orthogonality of the pair [39]. We seek to develop a strategy of switching the species specific identity and orthogonality of the previously evolved *M. jannaschii* nonnatural tRNA/aaRS pairs. This approach can be used to easily develop more mammalian useable pairs and incorporate more types of nonnatural amino acids. To our best of our knowledge, there has not been an engineered aaRS with altered specificity for both its tRNA and amino acid substrates. Our methodology would change the target specificity of *M. jannaschii* tyrosyl-tRNA synthetase so that it could aminoacylate a non-cognate tRNA with a nonnatural amino acid *in vivo.* The following specific aims will detail our approach in the design and creation of a nonnatural pair.

1.4.2 Specific Aim 1: Design and express a mammalian-orthogonal amber codon suppressing tRNA

Archaea and *eukaryotic* tyrosyl-tRNA and tyrosyl-tRNA synthetases are orthogonal to *E. coli* pairs but not to one another [24, 40]. Schimmel's group has made pioneering discoveries in how acceptor stem sequences can determine the amino acid and species identity of a tRNAs [40, 41]. We seek to engineer an amber codon suppressing tRNA, derived from *M. jannaschii* tyrosyl-tRNA that is orthogonal in mammalian cells. Various designs with mutations in the anticodon and acceptor stem regions will be evaluated based on orthogonality and functionality (suppression efficiency). Different plasmid designs will also be tested to identify the most optimal method for expressing *archaea* derived tRNA genes in mammalian cells.

1.4.3 Specific Aim 2: Engineer a mutant *M. jannaschii* tyrosyl-tRNA synthetase that can charge the tRNA from Specific Aim 1

An aminoacyl-tRNA synthetase must recognize two specific targets when carrying out the aminoacylation reaction: its cognate tRNA and the corresponding amino acid. Previous work by other groups and ourselves has shown that the specificity for one can be change without affecting the other [40, 42, 43]. We used rationale design in order to engineer a mutant synthetase derived from *M. jannaschii* tyrosyl-tRNA synthetase that could aminoacylate our orthogonal tRNA (**Specific Aim 1**) in a mammalian cell. Various synthetases designs are evaluated based on their aminoacylation activity and optimized for expression in mammalian cells. We investigate whether peptide transplantations can be used to switch an aaRS's tRNA specificity without eliminating its activity.

1.4.4 Specific Aim 3: Apply the method from Specific Aim 2 to a nonnatural aaRS

We hypothesize that the tRNA specificity of a mutant *M. jannaschii* tyrosyl-tRNA synthetase can be altered without affecting its amino acid specificity. This would allow us to shuttle many *M. jannaschii* tyrosyl-tRNA synthetase derived nonnatural aaRS that have been used to incorporate nonnatural amino acids in proteins in *E. coli* cells. We will use the method developed in Specific Aim 2 to change the orthogonality of a nonnatural aaRS. As a result, it should ignore endogenous mammalian tRNAs while recognizing the tRNA from Specific Aim 1. These two components complete the orthogonal suppressor tRNA/nonnatural aaRS pair that is required for site-specific incorporation. 3,4-dihydroxy-L-phenylalanine is a nonnatural amino acid that has successfully been incorporated in *E. coli* [26]. A corresponding nonnatural aaRS, evolved from *M. jannaschii* tyrosyl-tRNA synthetase, can recognize 3,4-dihydroxy-L-phenylalanine while discriminating against tyrosine. We will test the CP1 substitution on the 3,4-dihydroxy-L-phenylalanine aaRS and others in order to demonstrate our methodology of shuttling other nonnatural aaRS for use in mammalian cells.

1.5 REFERENCES

1. Wang, L. and P.G. Schultz, *Expanding the genetic code.* Angewandte Chemie-International Edition, 2005. **44**(1): p. 34-66.
2. Wold, F., *In vivo chemical modification of proteins (post-translational modification).* Annu Rev Biochem, 1981. **50**: p. 783-814.
3. Link, A.J., M.L. Mock, and D.A. Tirrell, *Non-canonical amino acids in protein engineering.* Current Opinion in Biotechnology, 2003. **14**(6): p. 603-609.
4. Heim, R. and R.Y. Tsien, *Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.* Curr Biol, 1996. **6**(2): p. 178-82.
5. Dawson, P.E., et al., *Synthesis of proteins by native chemical ligation.* Science, 1994. **266**(5186): p. 776-9.
6. Johnson, J.A., et al., *Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications.* Curr Opin Chem Biol. **14**(6): p. 774-80.
7. Calendar, R. and P. Berg, *The catalytic properties of tyrosyl ribonucleic acid synthetases from Escherichia coli and Bacillus subtilis.* Biochemistry, 1966. **5**(5): p. 1690-5.
8. Cohen, G.N. and R. Munier, *[Incorporation of structural analogues of amino acids in bacterial proteins].* Biochim Biophys Acta, 1956. **21**(3): p. 592-3.
9. Cowie, D.B., et al., *Amino acid analog incorporation into bacterial proteins.* Biochim Biophys Acta, 1959. **34**: p. 39-46.

10. Dieterich, D.C., et al., *Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT).* Proc Natl Acad Sci U S A, 2006. **103**(25): p. 9482-7.

11. Beatty, K.E. and D.A. Tirrell, *Two-color labeling of temporally defined protein populations in mammalian cells.* Bioorg Med Chem Lett, 2008. **18**(22): p. 5995-9.

12. Tcherkezian, J., et al., *Transmembrane receptor DCC associates with protein synthesis machinery and regulates translation.* Cell. **141**(4): p. 632-44.

13. Shoulders, M.D., et al., *Stereoelectronic and steric effects in side chains preorganize a protein main chain.* Proc Natl Acad Sci U S A. **107**(2): p. 559-64.

14. Yoo, T.H., A.J. Link, and D.A. Tirrell, *Evolution of a fluorinated green fluorescent protein.* Proc Natl Acad Sci U S A, 2007. **104**(35): p. 13887-90.

15. Lepthien, S., et al., *Azatryptophans endow proteins with intrinsic blue fluorescence.* Proc Natl Acad Sci U S A, 2008. **105**(42): p. 16095-100.

16. Kurschus, F.C., et al., *Gold fluorescent annexin A5 as a novel apoptosis detection tool.* Cytometry A, 2009. **75**(7): p. 626-33.

17. Budisa, N., et al., *Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: structure and stability of the per-thiaproline mutant of annexin V.* Proc Natl Acad Sci U S A, 1998. **95**(2): p. 455-9.

18. Ibba, M. and H. Hennecke, *Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows in vitro and in vivo synthesis of proteins containing unnatural amino acids.* FEBS Lett, 1995. **364**(3): p. 272-5.

19. Noren, C.J., et al., *A general method for site-specific incorporation of unnatural amino acids into proteins.* Science, 1989. **244**(4901): p. 182-8.

20. Bain, J.D., et al., *Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide.* Journal of the American Chemical Society, 1989. **111**(20): p. 8013-8014.

21. Goodman, H.M., et al., *Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA.* Nature, 1968. **217**(5133): p. 1019-24.

22. Nowak, M.W., et al., *In vivo incorporation of unnatural amino acids into ion channels in Xenopus oocyte expression system.* Methods Enzymol, 1998. **293**: p. 504-29.

23. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.

24. Wang, L., et al., *A New Functional Suppressor tRNA/Aminoacylâˆ’tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins.* Journal of the American Chemical Society, 2000. **122**(20): p. 5010-5011.

25. Wang, L., et al., *Addition of the keto functional group to the genetic code of Escherichia coli.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 56-61.

26. Alfonta, L., et al., *Site-specific incorporation of a redox-active amino acid into proteins.* J Am Chem Soc, 2003. **125**(48): p. 14662-3.

27. Umeda, A., et al., *Site-specific Protein Cross-Linking with Genetically Incorporated 3,4-Dihydroxy-L-Phenylalanine.* Chembiochem, 2009. **10**(8): p. 1302-1304.

28. Deiters, A., et al., *Site-specific PEGylation of proteins containing unnatural amino acids.* Bioorg Med Chem Lett, 2004. **14**(23): p. 5743-5.

29. Jevsevar, S., M. Kunstelj, and V.G. Porekar, *PEGylation of therapeutic proteins.* Biotechnol J. **5**(1): p. 113-28.

30. Bose, M., et al., *The incorporation of a photoisomerizable amino acid into proteins in E. coli.* J Am Chem Soc, 2006. **128**(2): p. 388-9.

31. Xie, J., et al., *The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination.* Nat Biotechnol, 2004. **22**(10): p. 1297-301.

32. Sakamoto, K., et al., *Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells.* Nucleic Acids Res, 2002. **30**(21): p. 4692-9.

33. Chin, J.W., et al., *An expanded eukaryotic genetic code.* Science, 2003. **301**(5635): p. 964-7.

34. Wang, W.Y., et al., *Genetically encoding unnatural amino acids for cellular and neuronal studies.* Nature Neuroscience, 2007. **10**(8): p. 1063-1072.

35. Ye, S., et al., *Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis.* J Biol Chem, 2008. **283**(3): p. 1525-33.

36. Zhang, Z., et al., *Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells.* Proc Natl Acad Sci U S A, 2004. **101**(24): p. 8882-7.

37. Liu, W., et al., *Genetic incorporation of unnatural amino acids into proteins in mammalian cells.* Nat Methods, 2007. **4**(3): p. 239-44.

38. Zhang, Z., et al., *A new strategy for the site-specific modification of proteins in vivo.* Biochemistry, 2003. **42**(22): p. 6735-46.

39. Steer, B.A. and P. Schimmel, *Major anticodon-binding region missing from an archaebacterial tRNA synthetase.* J Biol Chem, 1999. **274**(50): p. 35601-6.

40. Wakasugi, K., et al., *Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant.* EMBO J, 1998. **17**(1): p. 297-305.

41. Schimmel, P., et al., *An operational RNA code for amino acids and possible relationship to genetic code.* Proc Natl Acad Sci U S A, 1993. **90**(19): p. 8763-8.

42. Neumann, H., A.L. Slusarczyk, and J.W. Chin, *De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs.* J Am Chem Soc. **132**(7): p. 2142-4.

43. Thibodeaux, G.N., et al., *Transforming a Pair of Orthogonal tRNA-aminoacyl-tRNA Synthetase from Archaea to Function in Mammalian Cells.* Plos One, 2010. **5**(6): p. -.

Chapter Two

Engineering a Mammalian-Orthogonal Suppressor tRNA

2.1 An Orthogonal Amber Codon Suppressing *M. jannaschii* tyrosyl-tRNA

2.1.1 Introduction

An orthogonal amber coding suppressing tRNA is a key component in Schultz's method of site-specific nonnatural amino acid incorporation [1]. A nonnatural aminoacyl-tRNA synthetase (aaRS) can then be used to charge the suppressor tRNA with a nonnatural amino acid [2]. Protein translation normally terminates when a stop codon is encountered on the mRNA transcript. However, an amino acid charged suppressor tRNA can interact with the ribosome complex and allow translation to proceed past the stop codon [3]. As a result, the attached amino acid is incorporated into the protein in response to the stop codon and downstream codons are also translated. This chapter will focus on the design and expression of a tRNA that can suppress amber stop codons in mammalian cells. It must also be orthogonal to the host cell's aaRS. We assess a number of designs in order to identify the tRNA with the highest suppression efficiency and orthogonality. Chapters 3 and 4 will detail our efforts to create a nonnatural aaRS that can charge our suppressor tRNA with a nonnatural amino acid.

A tyrosyl-tRNA from *Methanococcus jannaschii* was chosen as the starting template for our suppressor tRNA because mutants of the *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) have been used extensively to incorporate nonnatural amino acids [4]. Several existing suppressor tRNAs have already been derived from *M. jannaschii* tyrosyl-tRNA but none of them are orthogonal to mammalian TyrRS [5]. This limits their use solely to *prokaryotic* systems. The first base pair in the tyrosyl-tRNA's acceptor stem has

been implicated to be a key determinant in whether aminoacylation can occur between tyrosyl-tRNAs/TyrRS from different species [6]. *Archaea* and *eukaryotic* tyrosyl-tRNAs contain a C1:G72 sequence and can be charged be each other's TyrRS. However, *prokaryotic* tyrosyl-tRNA acceptor stems have a G1:C72 sequence and is orthogonal to *archaea* and *eukaryotic* TyrRS (**Figure 2.1**). The importance of this pair is further demonstrated in works done by Quinn *et al.* [7] when it was shown that an *eukaryotic* TyrRS can no longer recognize and charge a tRNA microhelix after its acceptor stem was mutated from C1:G72 to G1:C72. Pioneering work by Schimmel and others have also investigated the importance of the acceptor stem sequence in determining the identity of a tRNA [8]. These studies suggest that the recognition of *M. jannaschii* tyrosyl-tRNA by *eukaryotic* TyrRS can be reduced by modifying its acceptor stem sequence.

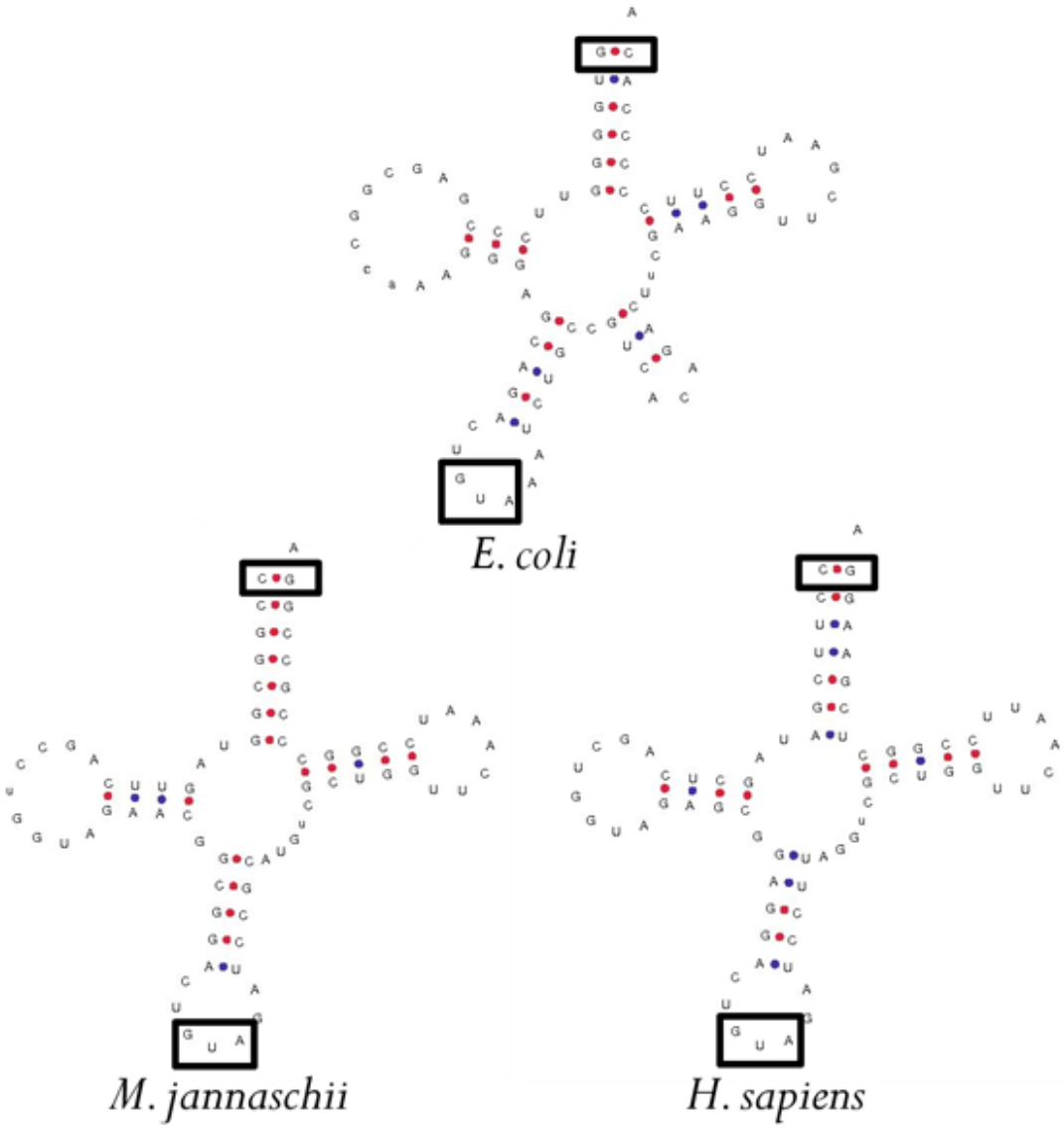


Figure 2.1: Secondary structure predictions of tyrosyl-tRNAs from *E. coli*, *M. jannaschii* and human. All three tRNAs contain a GUA sequence in their anticodon region. This allows them to correctly base pair with a TAC codon which corresponds to tyrosine in the genetic code. *E. coli* tyrosyl-tRNA contains a G1:C72 sequence in its acceptor stem region and is neither recognized by *M. jannaschii* nor human TyrRS. *M. jannaschii* and human tyrosyl-tRNA contain a C1:G72 sequence and can be charged by the each other's TyrRS. (tRNAscan-SE, UCSC)

2.1.2 Results and Discussion

A gene was first constructed to express a suppressor tRNA in HEK293T cells. Sequences from a human tyrosyl-tRNA gene were added to the 5' and 3' ends of a *M. jannaschii* tyrosyl-tRNA sequence. The anticodon region of the tRNA was then mutated from GTA to CTA which allows it to base pair with an amber stop codon (TAG). Six tandem-copies of this construct were cloned into pZeoSV2 (+) (Invitrogen, Carlsbad, CA) to make the plasmid 6x_wt-tRNA$_{CUA}$. An orthogonal suppressor tRNA was also made in a similar fashion except the acceptor stem contained a G1:72 base pair. Six tandem copies of this tRNA constructed were cloned into pZeoSV2 (+) to make the plasmid 6x_1bp-tRNA$_{CUA}$ (**Figure 2.2**).

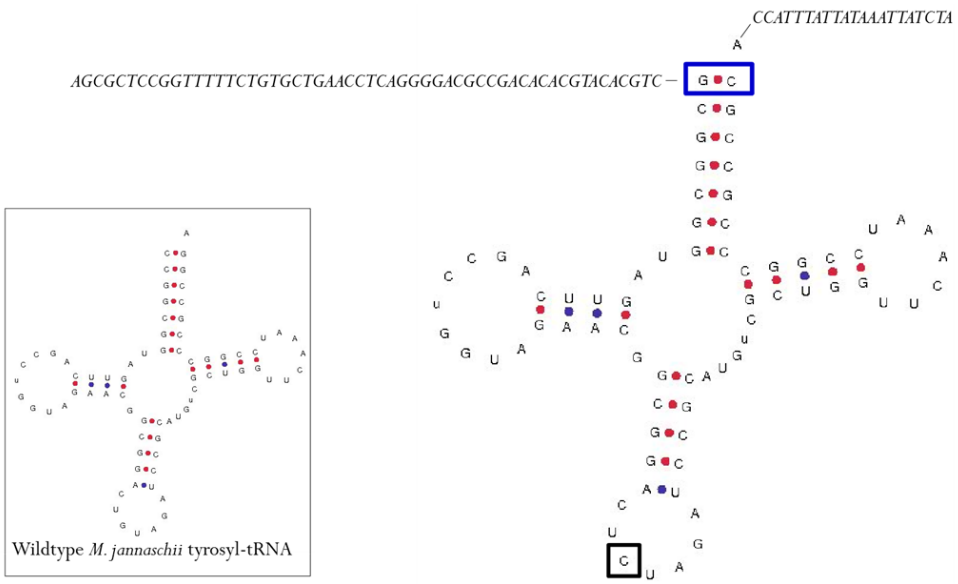


Figure 2.2: Sequence of the tRNA in the plasmid 6x_1bp-tRNA$_{CUA}$. A suppressor tRNA was first made by changing the anticodon from GUA to CUA. Swapping C1:C72 sequence in the acceptor stem with G1:C72 reduces its recognition by endogenous synthetases. Sequences were also added to the 5' and 3' ends of the gene to improve its transcription in HEK293T cells.

To test the suppression efficiency of the tRNA we introduced an amber stop codon into a T4 fibritin (foldon) gene at the 68th position. This mutant gene was then cloned into pCDNA3.1 (Invitrogen, Carlsbad, CA) which also contained a sequence for a V5 peptide on the C-terminus. HEK293T cells were transfected with a tRNA plasmid and the foldon_68TAG plasmid. Cells were harvested 48 hours later and full length foldon protein was purified and detected using anti-V5 antibodies (**Figure 2.4**). Codons after the TAG should not be translated into amino acids since translation terminates when the stop codon is encountered by the ribosome complex. Full length foldon protein (with a V5 tag) can only be detected if the cell has a charged suppressor tRNA (**Figure 2.3**). Full length foldon was observed in cells that were transfected with the 6x_wt-tRNA$_{CUA}$ plasmid suggesting that the HEK293T TyrRS was able to charge it with tyrosine. This supports previous studies by others that have demonstrated the aminoacylation of *M. jannaschii* tyrosyl-tRNA by an *eukaryotic* TyrRS [5]. On the other hand, cells transfected with the 6x_1bp-tRNA$_{CUA}$ plasmid produced very little full length foldon. This suggests that a single base pair swap in the acceptor stem was able to dramatically reduce the recognition of the tRNA by the host cell's TyrRS.

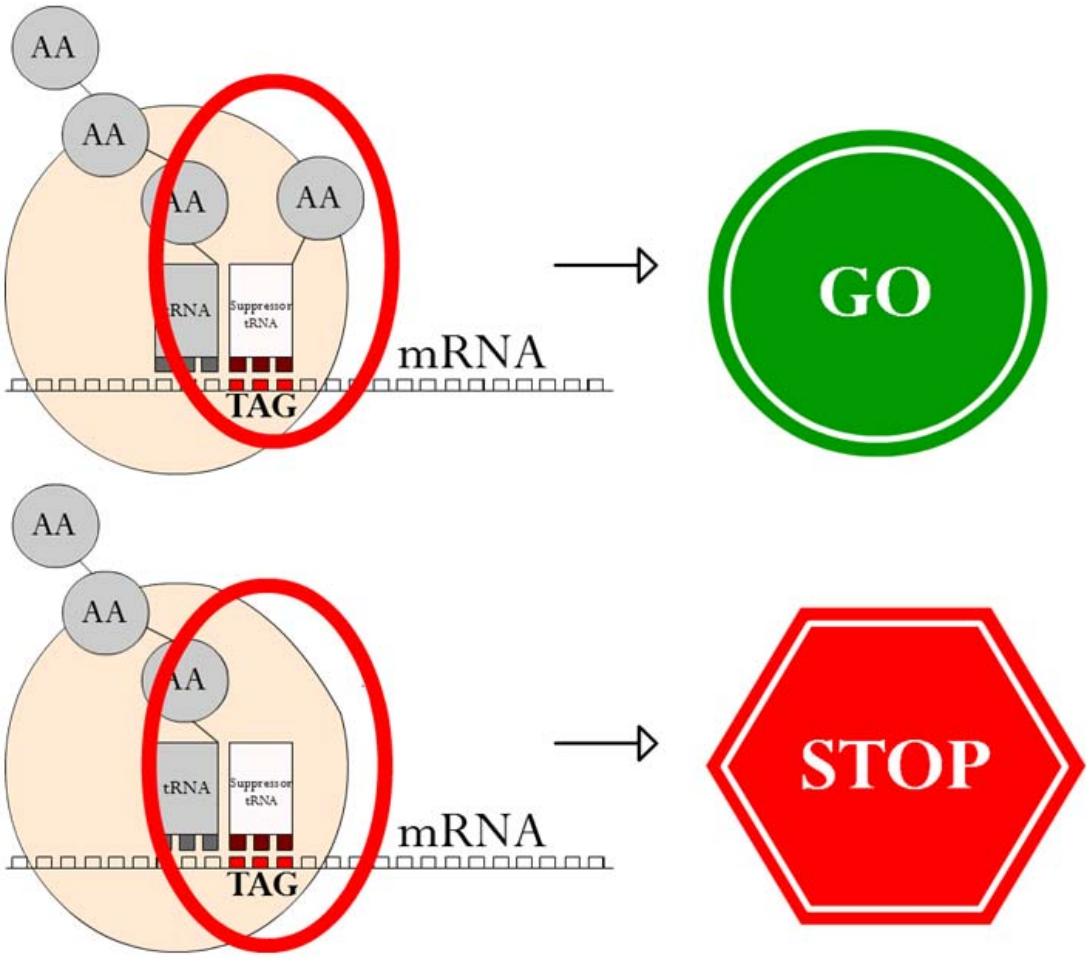


Figure 2.3: Translation of the mRNA transcription usually terminates when a stop codon (i.e. TAG/UAG) is encountered. Only a charged suppressor tRNA can be used to 'suppress' the stop codon and allow translation to continue.

Figure 2.4: Suppression of an amber stop codon in a foldon gene. Expression of full length foldon was probed using anti-V5 antibodies that are present on the C-terminal end if it is completely translated. Co-transfection of HEK293T cells with plasmids containing the tRNA and foldon_68TAG plasmids only resulted in full length foldon expression when the tRNA contained a C1:G72 sequence [9].

Green fluorescent protein (GFP) was also used as a reporter protein since it can be detected with UV-light. An amber stop codon (TAG) was introduced into the 40th position of a GFP gene and cloned into a plasmid. HEK293T cells were co-transfected with the GFP_40TAG and 6x_1bp-tRNA$_{CUA}$ plasmids. Fluorescent activated cell sorting (FACS) was used to quantify the amount of full length GFP expression. Introduction of the TAG into the GFP gene produced only a truncated GFP that was not detectable by FACS. Co-transfection with the 6x_1bp-tRNA$_{CUA}$ also restored < 1% (**Figure 3.5**) of full length GFP expression indicating that it was not charged by the host's TyrRS. This agrees with the outcomes of the foldon suppression assay. Results from this section demonstrate that an orthogonal amber codon suppressing tRNA can be created from *M. jannaschii* tyrosyl-tRNA by manipulating its acceptor stem sequences.

2.1.3 Materials and Methods

Cell Culturing and Transfections

Plasmids were amplified by transforming electrically component *E. coli* TOP10 cells, grown in LB, and isolated using QIAprep Spin Miniprep Kit (Qiagen, Madison, WI). HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2. Transfections were performed according to product manual using FuGENE 6 (Roche, Indianapolis, IN). Full length GFP was excited at 460-500 nm and detected using a Nikon Eclipse TE2000-S microscope equipped with a FITC HyQ filter (Chroma).

Plasmid Construction

A mutated *M. jannaschii* tyrosyl-tRNA gene containing changes in the anticodon (G38) and acceptor stem (G1:C72) regions was created from overlapping oligonucleotides

5'-AACTGCAGAGCGCTCCGGTTTTTCTGTGCTGAACCTCAGGGGACGC
CGACACACGTACACGTCGCGGCGGTAGTTCAGCCTGGTAGAACGGCG
GACTCTAAATCCGCATG-3'

5'-CCGCTCGAGTAGATAATTTATAATAAATGGTGCGGCGGGCCGGATT
TGAACCAGCGACATGCGGATTTAGAGTCCGCCGTTCTACCA-3'

Also included in these oligonucleotides are the 5' flanking sequences

5'-AGCGCTCCGGTTTTTCTGTGCTGAACCTCAGGGGACGCCGACAC
ACGTACACGTC-3'

and 3' flanking sequences from a human tyrosyl-tRNA gene:

5'-TTTATTATAAATTATCTA-3'

The oligonucleotides were annealed then extended using a Klenow (New England Biolabs, Ipswich, MA) enzyme for 30 minutes, according to the product manual. The mutant tRNA sequence was amplified by PCR, digested overnight with restriction enzymes (NEB) PstI and XhoI. Then it was ligated using T4 DNA ligase (NEB) into a plasmid, pZeoSV2 (+) (Invitrogen, Carlsbad, CA), pre-digested with PstI and XhoI. This resulting plasmid (1xtRNA) was used as a template to create our six-tandem copy plasmid. To create the six-tandem copy plasmid, the mutant tRNA gene was first amplified by PCR using the primers

5’-GGGGTACCAGCGCTCCGGTTTTTCTGTG-3’

5’-GGGTACCTAGATAATTTATAATAAATGGTGCGGCGGG-3’

The PCR product was then cleaned using a QIAprep Kit (Qiagen, Madison, WI), digested with restriction enzymes KpnI and ligated with T4 DNA ligase into a 1xtRNA plasmid predigested with KpnI creating a 2xtRNA plasmid. Another copy of the mutant tRNA gene was amplified by PCR using the primers

5’-GGAATTCAGCGCTCCGGTTTTTCTGTG-3’

5’-GGGAAGCTTGGGCCCTCTAGACTCGAG-3’

The resulting PCR product was cleaned, digested with restriction enzymes EcoRI. After purification, the product was ligated using T4 DNA ligase into a 2xtRNA plasmid pre-digested with EcoRI to create a 3xtRNA plasmid. Finally, to create the six-tandem copy plasmid three-tandem copies of the mutant tRNA gene were amplified by PCR using the primers:

5’-GGGGCTAGCTTAAGCTTGGTACCAGCGC-3’

5’-GGGAAGCTTGGGCCCTCTAGACTCGAG-3’

The resulting product was purified, digested with the restriction enzymes NheI and HindIII, purified and then ligated using T4 DNA ligase into a 3xtRNA plasmid pre-

digested with NheI and HindIII. The final 6xtRNA construct was confirmed by DNA sequencing (University of Texas Core Facility).

The 40th codon of a green fluorescent protein (GFP) gene under control of a CMV promoter in a mammalian expression vector (pLit) was mutated to an amber codon (TAG) using the Quikchange II Site-Directed Mutagenesis Kit (Stratagene) according to the product manual. The identity of the resulting GFP-40TAG plasmid was confirmed by DNA sequencing (University of Texas Core Facility).

The foldon-68TAG plasmid, which was used to express the bacteriophage T4 fibritin (foldon) domain in HEK293T cells, was constructed by inserting the PCR-amplified gene fragment into the pCDA3.1-V5-His-TOPO vector (Invitrogen, Carlsbad, CA). Plasmid foldon-68TAG, which contains an amber stop codon in its 68th position, was constructed by site directed mutagenesis by using the QuikChangeXL (Stratagene,) method and the corresponding HPLC-purified primers.

Harvest and Lysis of HEK293T Cells

72 hours after transfection, HEK293T cells were detached by physical scrapping and collected by centrifugation at 1000xg. A passive lysis buffer (Promega) supplemented with a complete protease cocktail inhibitor (Roche, Indianapolis, IN) was used to lyse the collected cells on ice for 30 minutes. A soluble protein fraction was collected from the centrifugation at 14,000 rpm for 20 minutes and subjected to western blot analysis.

Western Blots

The soluble protein fractions were separated by 12% SDS-PAGE and transferred onto an Immobilon PVDF membrane (Millipore) in a transfer buffer (48 mM Tris, 39

mM glycine, 20% methanol (v/v), pH 9.2) using a BioRad Semi-Dry Blotter. The membrane was then probed for full length foldon expression using an anti-V5 antibody (Invitrogen, Carlsbad, CA) and a goat anti-mouse alkaline phosphatase conjugated secondary antibody (Bio-Rad, Hercules, CA). The bands were visualized using a chemiluminscent reagent, PhosphaGlo AP (KPL) on a BioMax Light Film (Eastman Kodak Co., Rochester, NJ).

FACS Analysis

A BD FACSCalibur flow cytometer was used to gate 10,000 cells based on forward and side scatter. Each cell was then excited with 488 nm light and its resulting emission detected with an FL-1 filter (515-545 nm). Positive fluorescent populations were gated based on negative controls and analyzed using the computer software Cyflogic v.1.2.1 (CyFlo, LTD., Turku, Finland).

2.2 EXPRESSING AN *ARCHAEA* TRNA GENE WITH THE HUMAN H1 PROMOTER

2.2.1 Introduction

Our next focus was on improving the transcription efficiency of the *M. jannaschii* tRNA gene in HEK293T cells. Schultz's method of site-specific nonnatural amino incorporation is limited by the amount aminoacylated suppressor tRNA present in the cell. Our previous design consisted of multiple copies of the *archaea* tRNA gene flanked by sequences taken from a human tyrosyl-tRNA gene [9]. However, the transcriptional regulation of tRNA genes differs greatly between *eukaryotic* and *archaea* species. Sequences within *eukaryotic* tRNA genes, such as internal promoters, are usually required in order to initiate transcription [10]. A single type of RNA polymerase is used to transcribe both mRNA and tRNA in *archaea* cells while *eukaryotic* cells uses a different RNA polymerase for each [11]. Another key difference occurs in the post-transcriptional processing of the tRNA genes. The tRNA gene that is contained in 6x_1bp-$tRNA_{CUA}$ included a 3'CCA sequence but it has been shown that this is added post-transcriptional to the tRNA in *eukaryotic* cells [12]. All these factors suggest that our previous design was an inefficient way of expressing an *archaea* tRNA gene.

Wang *et al*. recently had success in using the human H1 promoter [13] to express an *E. coli* tyrosyl-tRNA gene in mammalian cells [14]. We reason that a similar approach can be used to express our *M. jannaschii* suppressor tRNA gene. A sequence of about 200 nucleotides, taken from the human H1 RNA gene, is sufficient in size to recruit RNA polymerase III and initiate transcription [13]. We constructed an entirely new plasmid consisting only of the human H1 promoter and a downstream tRNA gene with an added 3' flanking sequence. The multiple-tandem copies design was replaced in favor of a single tRNA gene, lacking the 3' CCA (**Figure 2.5**). We relied on endogenous post-transcriptional mechanisms to add this sequence. Using this plasmid we tested its ability

to express functional tRNA by using it to express a non-orthogonal suppressor tRNA (wt-$tRNA_{CUA}$).

2.2.2 Results and Discussion

We designed a suppression assay to test the ability of the human H1 plasmid design to express tRNA genes. A gene for *M. jannaschii* tyrosyl-tRNA was first cloned downstream of the human H1 promoter (**Figure 2.5**). The anticodon region of this tRNA was modified from GUA to CUA to make H1_wt-$tRNA_{CUA}$. Another plasmid was used to harbor and express a gene for green fluorescent protein (GFP) that contained an amber stop codon in the 40^{th} position (GFP_40TAG). This is similar to the suppression assays discussed in the previous section when we calculate suppression efficiency based on the expression of full length protein. The total amount of full length GFP was detected by UV-light and quantified using fluorescent activated cell sorting (FACS). If the tRNA was non-orthogonal then it should be easily aminoacylated by endogenous TyrRS. The stop codon in the GFP gene would only be suppressed by an aminoacylated suppressor tRNA, resulting in the expression of full length GFP.

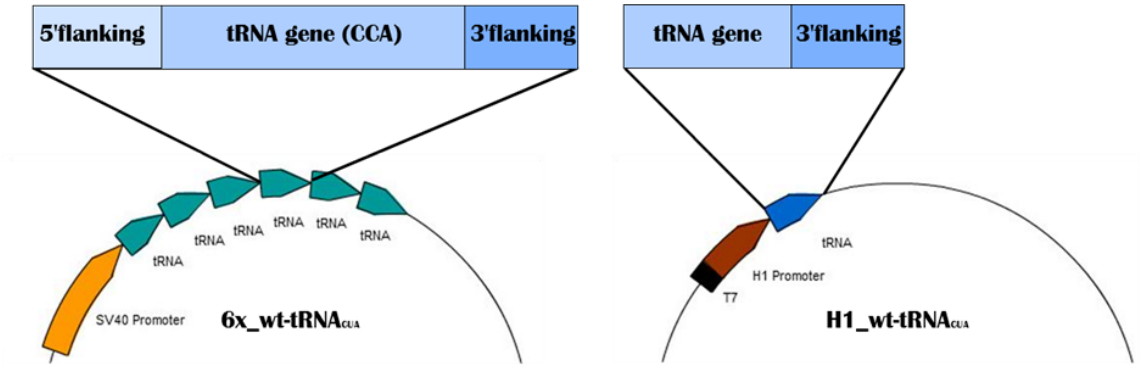


Figure 2.5: Comparing the 6x_wt-tRNA$_{CUA}$ and H1-wt-tRNA$_{CUA}$ plasmid constructions. The six-tandem approach is discarded in favor of an upstream human H1 promoter. A 3' flanking sequence was also used to terminate transcription. tRNA gene sequences are identical except in the 3' CCA tail. 6x_wt-tRNA$_{CUA}$ contains the sequence but H1_wt-tRNA$_{CUA}$ does not.

HEK293T cells were transfected with the GFP_40TAG and H1_wt-tRNA$_{CUA}$ plasmids. A plasmid harboring a wild type GFP gene was also used as a positive control. GFP expression was monitored using UV-light and quantified 72 hours after transfection using FACS analysis (**Figure 2.6**). Addition of the amber stop codon into the GFP gene resulted in almost no full length GFP expression. When H1_wt-tRNA$_{CUA}$ was present in the cell, it was recognized and charged by HEK293T TyrRS. As a result, a functional suppressor tRNA was generated that was able to restore a third of the full length GFP expression. Finally, to show that the observed fluorescence was a result of GFP expression we used anti-GFP antibodies to confirm its identity (**Figure 2.7**). The results of our studies show that the human H1 promoter can be used to transcribe high levels of an *archaea* tRNA gene. This method should allow us to produce more nonnatural amino acid incorporated proteins by increasing the amount of nonnatural amino acid charged tRNA.

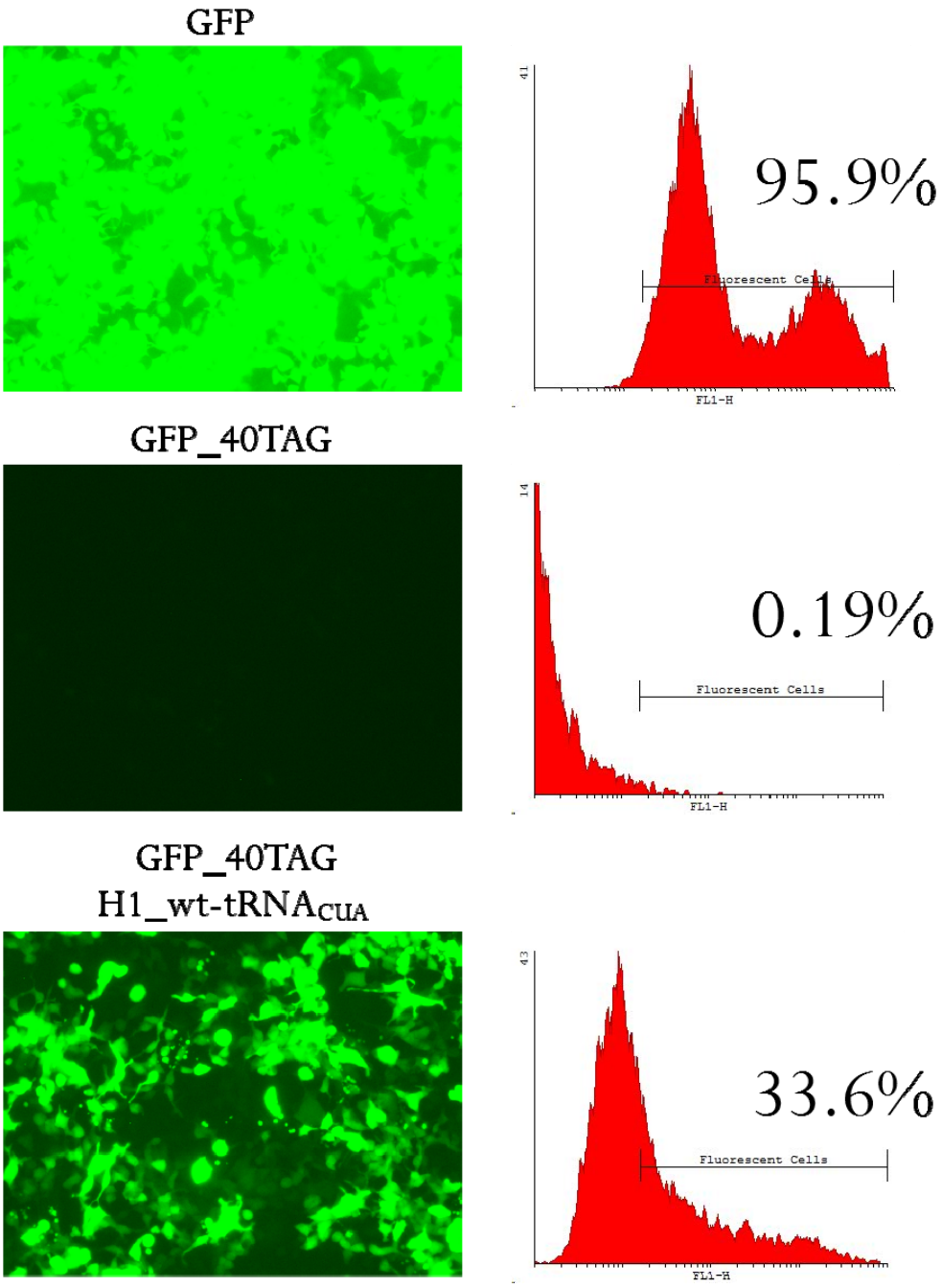


Figure 2.6: Almost no full length GFP can be detected after an amber stop codon is introduced into the gene. When wt-$tRNA_{CUA}$ is also present in the cell, endogenous TyrRS can recognize and charge it with tyrosine. This creates a functional suppressor tRNA that can be used to suppress the amber stop codon resulting in full length GFP. Our *M. jannaschii* tyrosyl-tRNA derived gene can be efficiently expressed using the H1 promoter. (Cyflogic)

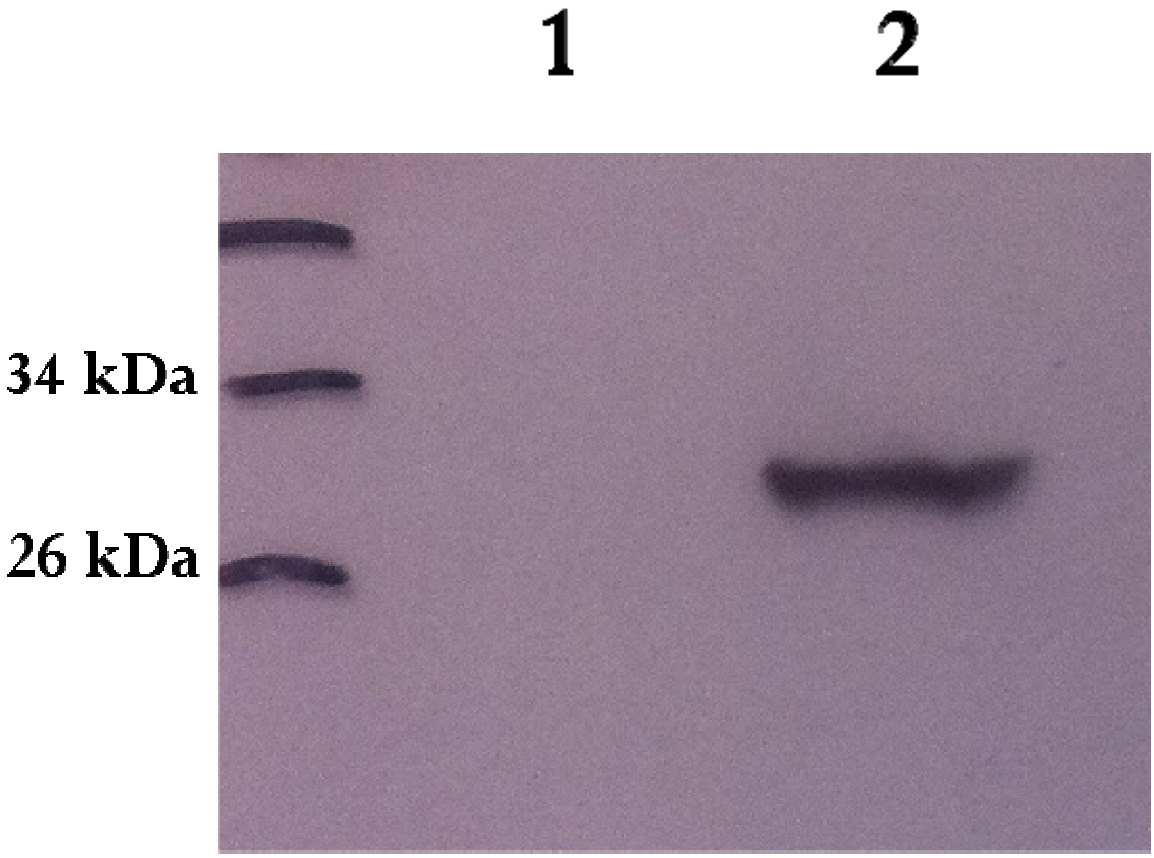


Figure 2.7: Detection of full length GFP with anti-GFP antibodies. HEK293T cells were transfected and harvested 72 hours later. (**1**) GFP_40TAG (**2**) GFP_40TAG and H1_wt-tRNA$_{CUA}$. A band that corresponded to the molecular weight of GFP (~27 kDa) was detected in cells transfected with the nonnatural suppressor tRNA.

2.2.3 Materials and Methods

Cell Culturing and Transfections

All plasmids were amplified in *E. coli* TOP10 cells and isolated using Qiagen's Miniprep Kits. HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2.

Transfections were performed according to product manual using FuGENE HD (Roche, Indianapolis, IN). Full length GFP was excited at 460-500 nm and detected using a Nikon Eclipse TE2000-S microscope equipped with a FITC HyQ filter (Chroma).

Construction of the GFP_40TAG Plasmid

A commercially available plasmid p-EGFPN1 (Clontech) was used to express a gene for EGFP using a CMV promoter. The 40th codon in the EGFP gene was mutated from TAC to TAG using site-directed mutagenesis (SDM) using primers designed by Stratagene's online primer design program

5'-GCGAGGGCGAGGGCGATTAGACCTACGGCAAGC-3'

5'-GCTTGCCGTAGGTCTAATCGCCCTCGCCCTCGC-3'

The QuickChange II Site-Directed Mutagenesis Kit was purchased from Stratagene (Stratagene, Agilent Technologies, Inc., Santa Clara, CA), and site-directed mutagenesis was completed according to the provided protocol. Plasmids were sequenced by the University of Texas at Austin ICMB DNA Sequencing Facility. The new plasmid was termed GFP_40TAG.

Construction of the Human H1 Promoter tRNA Plasmid (H1_wt-tRNA$_{CUA}$)

The MCS-1 region from a commercially available plasmid, pTRE-TIGHT-BI (*Clontech*) was removed by digested the plasmid with restriction enzymes KpnI and EcoRI (NEB, Ipswich, MA). Oligonucleotides containing a sequence for the T7 promoter:

5'-GGGGTACCCCTAATACGACTCACTATAGGGGGAATTCC-3'

5'-GGAATT-CCCCCTATAGTGAGTCGTATTAGGGGTACCCC-3'

were annealed, digested with KpnI and EcoRI. The digested insert was then ligated using DNA T4 Ligase into the pre-digested p-TRE-TIGHT-BI to form p-TRE-TIGHT-T7. Next, oligonucleotides containing the sequence for the human H1 promoter were annealed and extended using a klenow enzyme.

5’-CAACCCGCTCCAAGGAATCGCGGGCCCAGTGTCACTAGGCGGGAAC
ACCCAGCGCGCGTGCGCCCTGGCAGGAAGATGGCTGTGAGGGACAGG
GGAGTGGCGCCCTGCAA-‘3

5’-GAACTTATAAGATTCCCAAATCCAAAGACATTTCACGTTTATGGTG
ATTTCCCAGAACACATAGCGACATGCAAATATTGCAGGGCGCCACTCC
CCTGTCCCTCACAGCC-‘3

The sequence was then amplified by PCR using the primers

5’-GGAATTCCAATTCGAACGCTGACGTCATCAACCCGCTCCAAGG
AATC-‘3

5’-GAAGATCTGTGGTCTCATACAGAACTTATAAGATTCCCA-‘3

The resulting product was digested with restriction enzymes KpnI and BglII and ligated into a pre-digested p-TRE-TIGHT-T7 plasmid to make pT7-H1.

FACS Analysis

A BD FACSCalibur flow cytometer was used to gate 10,000 cells based on forward and side scatter. Each cell was then excited with 488 nm light and its resulting emission detected with an FL-1 filter (515-545 nm). Positive fluorescent populations were gated based on negative controls and analyzed using the computer software Cyflogic.

2.3 Evaluating the Orthogonality of Various *M. jannaschii* tyrosyl-tRNA Derived Suppressor tRNAs

2.3.1 Introduction

Studies by Schimmel and others have identified an 'operational RNA code' that exists within the sequences of a tRNA's acceptor stem [8]. Aminoacyl-tRNA synthetases (aaRS) recognize the base pairs in these regions in order to identify their cognate tRNA. For example, Schimmel has shown that the first base pair in the acceptor stem of tyrosyl-tRNA determines whether cross-species aminoacylation occurs. *Archaea* and *eukaryotic* tyrosyl-tRNA synthetase (TyrRS) charge each other's tyrosyl-tRNA (C1:G72) but not *prokaryotic* tyrosyl-tRNA (G1:C72) [6]. We showed in a previous section that mutating *M. jannaschii* tyrosyl-tRNA's G1:C72 to C1:G72 allows it to evade recognition by *eukaryotic* TyrRS. *E. coli* tyrosyl-tRNA is orthogonal in mammalian cells and differs from *M. jannaschii* tyrosyl-tRNA in the first three base pairs of its acceptor stem. In this section we investigate the effects on orthogonality by modifying additional pairs within the *M. jannaschii* tyrosyl-tRNA's acceptor stem. The CP1 region has been shown to directly interact with these sequences and additional mutations could affect the folding and structure in that region. The crystal structure of the *M. jannaschii* tyrosyl-tRNA in complex with its cognate TyrRS also shown that the TyrRS interacts directly with other sequences in the acceptor stem in addition to the C1:G72 pair [15].

The ability to rapidly test the orthogonality of various tRNA constructs is a huge benefit in using the human H1 promoter plasmid to express a tRNA gene. PCR and cloning steps are greatly reduced since only one copy of the tRNA gene was required as opposed to six or more copies. Using this plasmid, we tested the effects on orthogonality of additional acceptor stem modifications.

2.3.2 Results and Discussion

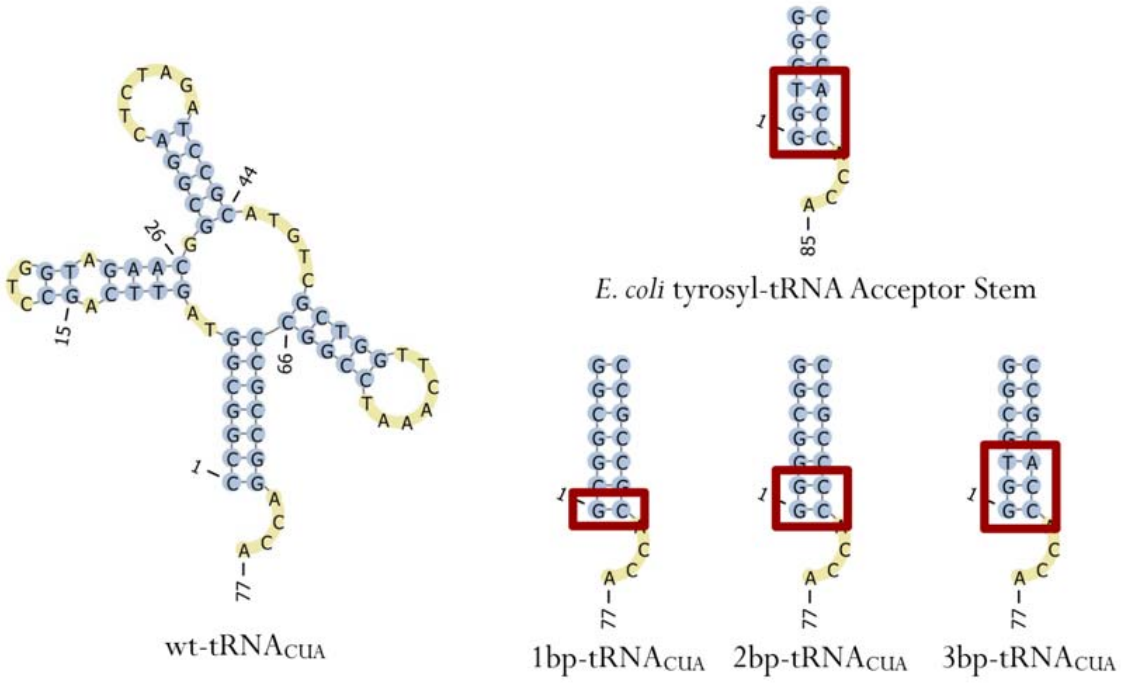


Figure 2.8: Comparing the acceptor stem sequences of various tRNA constructs. Wt-$tRNA_{CUA}$ is known to be non-orthogonal while *E. coli* tyrosyl-tRNA is orthogonal. Each tRNA gene was cloned into a plasmid, downstream of the human H1 promoter. We use a suppression assay to evaluate each design based on its recognition by HEK293T TyrRS (pknotsRG, BiBiServ)

Once again, a suppression assay was used to measure the orthogonality of each tRNA design. We co-transfected HEK293T cells with a mammalian expression plasmid that contained a gene for GFP_40TAG and a plasmid containing one of four tRNA designs (**Figure 2.8**). The tRNA plasmid was constructed as described in the previous section: an upstream human H1 promoter followed by a tRNA gene which has its 3'CCA tail replaced with a 3' flanking sequence from a human tyrosyl-tRNA gene. A single plasmid transfection with a mammalian expression plasmid containing a gene for GFP was also used as a positive control. And a single plasmid transfection with a gene for GFP_40TAG was used as a negative control.

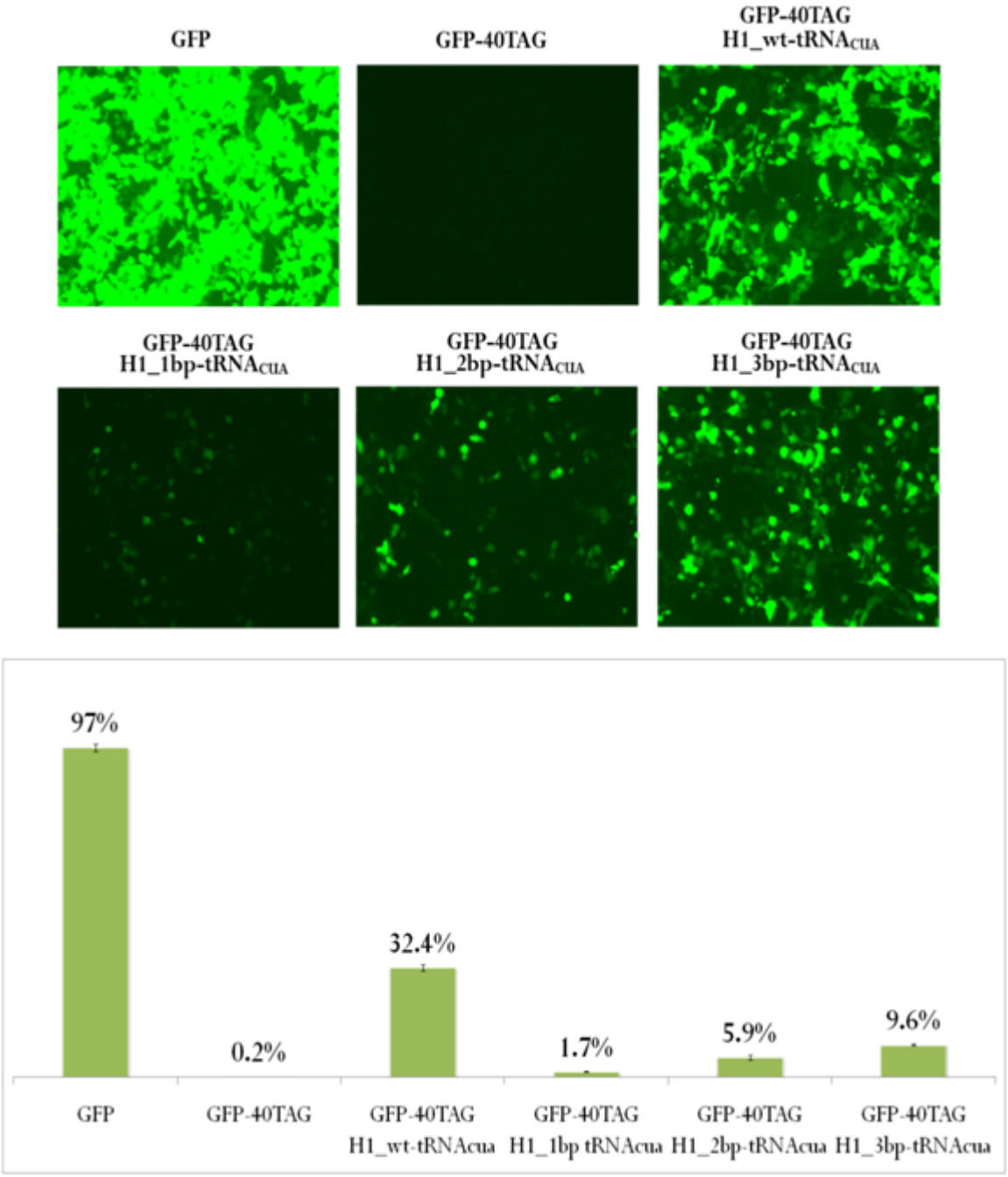


Figure 2.9: Assessing the orthogonality of various tRNA constructs in HEK293T cells. 72 hours after transfection FACS was used to quantify population of fluorescent cells. Wt-tRNA$_{CUA}$ was non-orthogonal as expected. However both 2bp-tRNA$_{CUA}$ and 3bp-tRNA$_{CUA}$ were less orthogonal than 1bp-tRNA$_{CUA}$ even though they had sequences that more closely matched *E. coli* tyrosyl-tRNA's acceptor stem.

Full length GFP was detected in HEK293T by UV-light and quantified using FACS (**Figure 2.9**). Large amounts of the protein are seen in cells after transfection with the GFP plasmid. This is dramatically reduced when an amber stop codon (TAG) was introduced into the gene. Transcription of the GFP_40TAG gene should still occur as normal but translation would halt once the ribosome complex encounters an amber stop codon. Since no full length GFP is produced, no fluorescence is observed. Co-transfections of HEK293T cells with the GFP_40TAG and H1_wt-tRNA$_{CUA}$ plasmids restored nearly a third of the full length GFP. Wt-tRNA$_{CUA}$ does not contain any acceptor stem modifications and thus is not orthogonal in mammalian cells as previously reported [5]. Endogenous aaRS in the HEK293T cells recognized and aminoacylated the suppressor tRNA. A change in the first base pair of the acceptor stem from C1:G72 to G1:C72 (1bp-tRNA$_{CUA}$) once again resulted in very low amounts of full length GFP (~**1.7%**). To our surprise, additional mutations in the acceptor stem sequence actually increased the expression of full length GFP. Previous work by other labs have demonstrated that *E. coli* tyrosyl-tRNA is orthogonal in mammalian cells [14]. Due to the important of sequences within the acceptor stem in determining tRNA identity, we expected 3bp-tRNA$_{CUA}$ to be the most orthogonal since the first four base pairs in its acceptor stem (G1:C72, G2:C71, T3:A70,G4:C69) are identical to *E. coli* tyrosyl-tRNA. However, both it and 2bp-tRNA$_{CUA}$ increased full length GFP expression by several magnitudes. A search of known human tRNA sequences revealed that two leucyl-tRNA genes also contain the same acceptor stem sequences as 3bp-tRNA$_{CUA}$ (G1:C72, G2:C71, U3:A70, and A73). Thus it could be possible that endogenous leucyl-tRNA synthetases are recognizing our tRNA and charging it with leucine. Further work will need to be done in this area.

Finally we wanted to test the functionality of the 1bp-tRNA$_{CUA}$ to ensure that it is folding properly *in vivo*. Low full length GFP expression infers that the suppressor tRNA was not aminoacylated by endogenous aaRS. Other than orthogonality, non-aminoacylation of the tRNA could also suggest that it is not functional or improperly folded. Therefore, we transfected HEK293T cells with the plasmids GFP_40TAG, H1_1bp-tRNA$_{CUA}$ and a mammalian expression plasmid containing the gene for an *E. coli* TyrRS. *E.coli* TyrRS has been shown to preferentially recognize and charge tyrosyl-tRNAs mini-helices containing the G1:C72 base pair [6]. If 1bp-tRNA$_{CUA}$ is orthogonal and functional, then the *E. coli* TyrRS should recognize and aminoacylate it, restoring the expression of full length GFP. 72 hours after transfection, cells were harvested and the population of fluorescing cells measured using FACS (**Figure 2.10**).

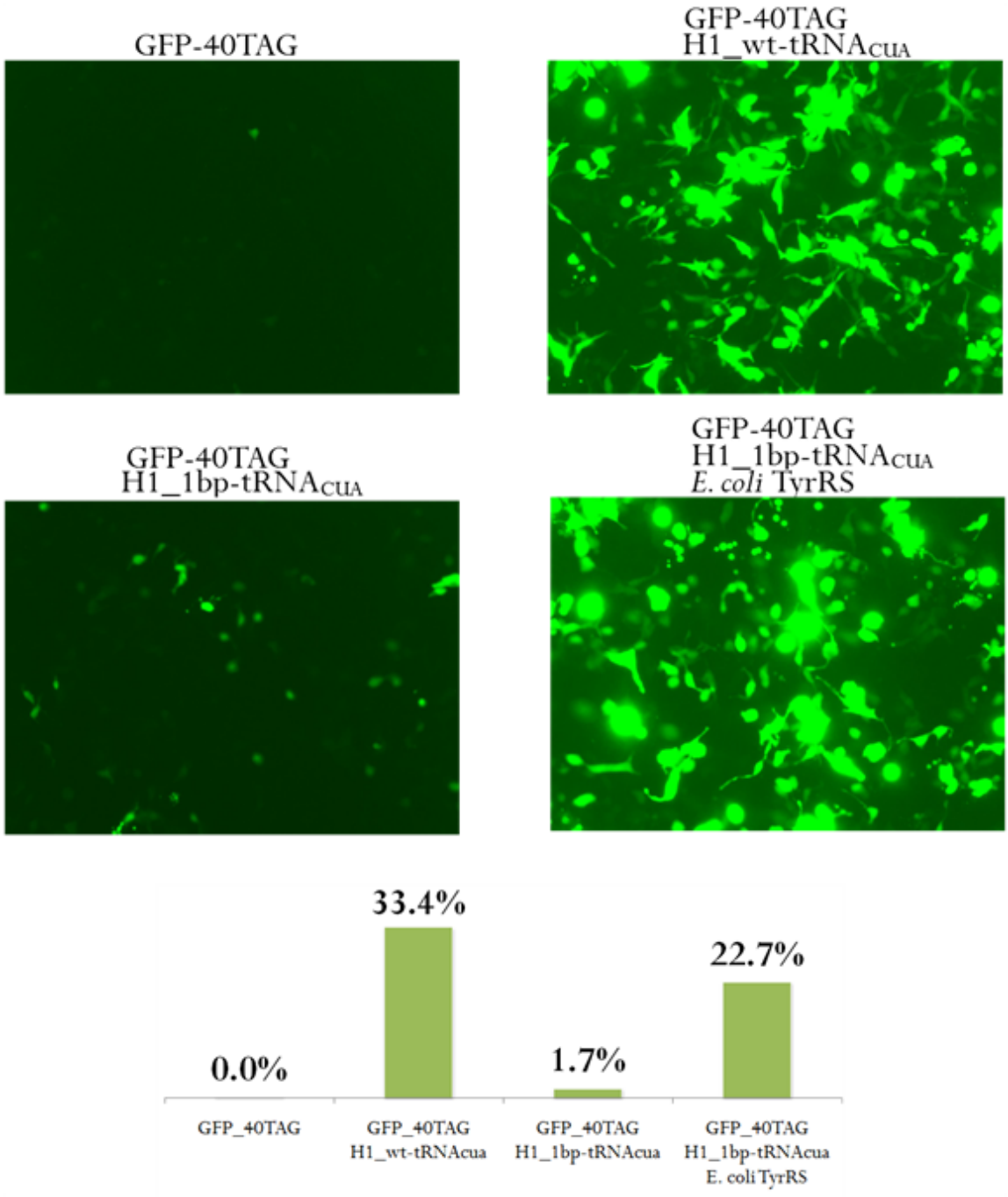


Figure 2.10: An amber stop codon in the GFP gene results in almost no detectable full length GFP. Only a charged suppressor tRNA can suppress the stop codon. 1bp-$tRNA_{CUA}$ cannot be recognized by mammalian TyrRS (orthogonal) but is charged by *E. coli* TyrRS. This demonstrates that 1bp-$tRNA_{CUA}$ is expressed, processed, and folded correctly.

The addition of *E. coli* TyrRS in HEK293T cells restored a significant amount of full length GFP expression. If 1bp-tRNA$_{CUA}$ was not functional, then we should expect to see a similar amount of full length GFP when *E. coli* TyrRS was present or absent. The non-orthogonal wt-tRNA$_{CUA}$ was chosen as a positive control since it represented the maximum amount of functional suppressor tRNA that could be generated by human H1 promoter plasmid. When *E. coli* TyrRS was present, the amount of full length GFP increased to was almost 75% of the control. This suggests that the 1bp-tRNA$_{CUA}$ was readily aminoacylated by *E. coli* TyrRS and then able to suppressor the amber stop codon in the GFP gene. Our investigations in this chapter support previous results that demonstrate the importance of the G1:C72 or C1:G72 in determining species specific identify of the tyrosyl-tRNA [6]. We were able to engineer an orthogonal amber codon suppressor tRNA for use in HEK293T cells. This completes the first component required for site-specific incorporation of nonnatural amino acids. The next two chapters will concentrate on our engineering of the corresponding nonnatural aaRS that can be paired with 1bp-tRNA$_{CUA}$.

2.3.4 Materials and Methods

Cell Culturing and Transfections

Plasmids were amplified by transforming electrically component *E. coli* TOP10 cells, grown in LB, and isolated using QIAprep Spin Miniprep Kit (Qiagen, Madison, WI). HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2. Transfections were performed according to product manual using FuGENE 6 (Roche,

Indianapolis, IN). Full length GFP was excited at 460-500 nm and detected using a Nikon Eclipse TE2000-S microscope equipped with a FITC HyQ filter (Chroma).

Construction of H1_1bp-tRNA$_{CUA}$, H1_2bp-tRNA$_{CUA}$, H1_3bp-tRNA$_{CUA}$

Two oligonucleotides for the 1bp-tRNA$_{CUA}$ gene was were annealed, extended using a klenow enzyme, and amplified using PCR. Sequence of oligonucleotides below:

5'-GAAGATCTGCGGCGGTAGTTCAGCCTGGTAGAACGGCGGACTCT
AAATCCGCATGTCGCTGGTTCAAATCCGGCCCGCCGCAGACAAGTGCG
GTTTTTTT-3

5'-CCAATGCATTGGTTGCCCGCTCGAGTAGAAAAAAACCGCACTTGTC
TGCGGCGGGCCGGATTTGAACCAGCGACATGCGGATTTAGAGTCCGCC
GTTCTA-3'

The PCR product was then digested with BglII and PstI and ligated into a pre-digested pT7-H1 (described above) plasmid to create 1bp-tRNA$_{CUA}$ plasmid.

H1_2bp-tRNA$_{CUA}$ was created in a similar fashion using the following oligonucleotides to create the tRNA gene:

5'-GAAGATCTGGGGCGGTAGTTCAGCCTGGTAGAACGGCGGACTCTA
AATCCGCATGTCGCTGGTTCAAATCCGGCCCGCCCCAGACAAGTGCGG
TTTTTTT-3'

5'-CCAATGCATTGGTTGCCCGCTCGAGTAGAAAAAAACCGCACTTGTC
TGGGGCGGGCCGGATTTGAACCAGCGACATGCGGATTTAGAGTCCGCC
GTTCTA-3

Finally, H1_3bp-tRNA$_{CUA}$ was made using the following oligonucleotides:

5'-GAAGATCTGGTGCGGTAGTTCAGCCTGGTAGAACGGCGGACTCTA
AATCCGCATGTCGCTGGTTCAAATCCGGCCCGCACCAGACAAGTGCGG
TTTTTTT-3’

5'-CCAATGCATTGGTTGCCCGCTCGAGTAGAAAAAAACCGCACTTG
TCTGGTGCGGGCCGGATTTGAACCAGCGACATGCGGATTTAGAGTCCG
CCGTTCTA-3'

Constructing the E. coli TyrRS Plasmid

A gene for *E. coli* TyrRS was amplified from a plasmid by PCR using the following primers:

5’-GCTCTAGATTATTTCCAGCAAATCAGACAGTA-3’

5’-CGGGATCCATGGCAAGCAGTAACTTGATTAAA-3’

The PCR product was digested with restriction enzymes XbaI and BamHI (New England Biolabs, Ipswich, MA) and ligated using T4 ligase into a pre-digested pEF6-V5 (Invitrogen, Carlsbad, CA) to express *E. coli* TyrRS in HEK293T cells.

FACS Analysis

A BD FACSCalibur flow cytometer was used to gate approximately 10,000 cells based on forward and side scatter. Each cell was then excited with 488 nm light and its resulting emission detected with an FL-1 filter (515-545 nm). Positive fluorescent populations were gated based on negative controls or positive controls and analyzed using the computer software Cyflogic v.1.2.1 (CyFlo, Ltd., Turku, Finland).

2.5 REFERENCES

1. Noren, C.J., et al., *A general method for site-specific incorporation of unnatural amino acids into proteins.* Science, 1989. **244**(4901): p. 182-8.
2. Liu, D.R., et al., *Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo.* Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10092-7.
3. Goodman, H.M., et al., *Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA.* Nature, 1968. **217**(5133): p. 1019-24.
4. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.
5. Steer, B.A. and P. Schimmel, *Major anticodon-binding region missing from an archaebacterial tRNA synthetase.* J Biol Chem, 1999. **274**(50): p. 35601-6.
6. Wakasugi, K., et al., *Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant.* EMBO J, 1998. **17**(1): p. 297-305.
7. Quinn, C.L., N. Tao, and P. Schimmel, *Species-specific microhelix aminoacylation by a eukaryotic pathogen tRNA synthetase dependent on a single base pair.* Biochemistry, 1995. **34**(39): p. 12489-95.
8. Schimmel, P., et al., *An operational RNA code for amino acids and possible relationship to genetic code.* Proc Natl Acad Sci U S A, 1993. **90**(19): p. 8763-8.

9. Thibodeaux, G.N., et al., *Transforming a Pair of Orthogonal tRNA-aminoacyl-tRNA Synthetase from Archaea to Function in Mammalian Cells.* Plos One, 2010. **5**(6): p. -.

10. Sharp, S., et al., *Internal control regions for transcription of eukaryotic tRNA genes.* Proc Natl Acad Sci U S A, 1981. **78**(11): p. 6657-61.

11. Weinmann, R., H.J. Raskas, and R.G. Roeder, *Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection.* Proc Natl Acad Sci U S A, 1974. **71**(9): p. 3426-39.

12. Deutscher, M.P., *Processing of tRNA in prokaryotes and eukaryotes.* CRC Crit Rev Biochem, 1984. **17**(1): p. 45-71.

13. Myslinski, E., et al., *An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene.* Nucleic Acids Res, 2001. **29**(12): p. 2502-9.

14. Wang, W.Y., et al., *Genetically encoding unnatural amino acids for cellular and neuronal studies.* Nature Neuroscience, 2007. **10**(8): p. 1063-1072.

15. Kobayashi, T., et al., *Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion.* Nat Struct Biol, 2003. **10**(6): p. 425-32.

Chapter Three

Altering the tRNA Specificity of an Aminoacyl-tRNA Synthetase

3.1 A CP1 Substituted *M. jannaschii* tyrosyl-tRNA Synthetase Charges 1bp-tRNA$_{CUA}$

3.1.1 Introduction

Chapter two concentrated on our efforts to design and express a functional amber codon suppressing tRNA (1bp-tRNA$_{CUA}$) that is orthogonal to mammalian aminoacyl-tRNA synthetases (aaRS). Now we shift our area of focus to the engineering of an aaRS that can charge 1bp-tRNA$_{CUA}$ with a nonnatural amino acid. These two components will form an orthogonal pair that can be used to genetically encode a nonnatural amino acid in response to the amber stop codon [1]. Several nonnatural aaRS have been evolved from the *M. jannaschii* tyrosyl-tRNA synthetase (TyrRS) and are capable of charging *M. jannaschii* tyrosyl-tRNA derived suppressor tRNAs with a nonnatural amino acid [2]. However, none of these nonnatural aaRS are not orthogonal in mammalian systems and will therefore aminoacylate endogenous tyrosyl-tRNAs. They also interact with a C1:G72 sequence in the acceptor stem region of their corresponding suppressor tRNA and would probably not recognize our 1bp-tRNA$_{CUA}$. In order to complete our pair we will have to alter the tRNA recognition of these nonnatural aaRS.

Schimmel and Wakasugi have previously demonstrated that the tRNA recognition of an aaRS can be changed without affecting its amino acid specificity. They constructed a mutant human TyrRS by substituting its connective polypeptide 1 (CP1) domain with the corresponding amino acids from the *Escherichia coli* TyrRS. The CP1 substituted human TyrRS was able to aminoacylate an *E. coli* tyrosyl-tRNA *in vitro* [3]. Their

seminal work showed that a TyrRS, which normal recognizes a C1:G72 tRNA can be manipulated into charging a G1:C72 tRNA.

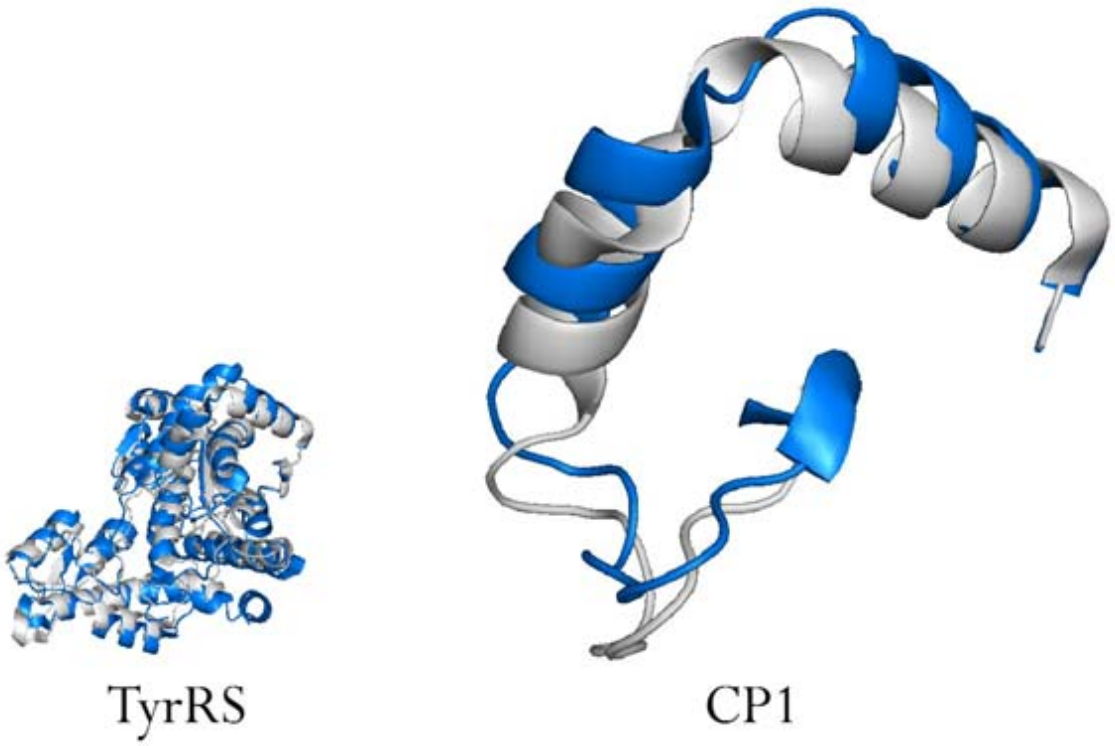


Figure 3.1: Crystal structure alignments of TyrRS from human (blue) and *M. jannaschii* (grey). The enzymes share a very similar 3-D structure and only charges C1:G72 containing tyrosyl-tRNAs. A CP1 substituted human TyrRS was able to charge an *E. coli* tyrosyl-tRNA which contains a G1:C72 acceptor stem sequence [3]. Alignments and images were created with the PyMOL Molecular Graphics System (pymol.org)

Both human and *M. jannaschii* TyrRS recognize tyrosyl-tRNAs containing a C1:G72 sequence while *prokaryotic* TyrRS recognize G1:C72. Human and *M. jannaschii* TyrRS also share a very similar 3-D crystal structure, especially in the CP1 domain (**Figure 3.1**). Strong structural and functional homology between the two TyrRS suggests that a CP1 substitution could also be used to manipulate the *M. jannaschii* TyrRS into

charging a G1:C72 containing tRNA (1bp-tRNA$_{CUA}$). The CP1 switched TyrRS should also have increased orthogonality in mammalian systems since its recognition of endogenous *eukaryotic* C1:G72 containing tyrosyl-tRNAs will be reduced. In this chapter we will investigate how changes in the CP1 region of *M. jannaschii* TyrRS affect its tRNA specificity. We later hope to apply this approach to some previously evolved nonnatural aaRS (evolved from *M. jannaschii* TyrRS). A functional nonnatural aaRS that could recognize our 1bp-tRNA$_{CUA}$ would complete the orthogonal pair and can be used to genetically incorporate a nonnatural amino acid in a mammalian cell (**Chapter 4).**

3.1.2 Results and Discussion

PyMOL Molecular Graphics System first was used to superimpose the crystal structures of the human (PDB: 1N3L) and *M. jannaschii* (PDB: 1J1U) TyrRS (**Figure 3.1**). Based on the similarity in structure, we identified a segment of the *M. jannaschii* TyrRS that corresponded to the human TyrRS's CP1 domain [3]. These amino acids were removed and replaced with an *E. coli* TyrRS CP1 sequence (**Figure 3.2**). The gene for this CP1 substituted *M. jannaschii* TyrRS was cloned into pEF6-V5-His6-TOPO (Invitrogen, Carlsbad, CA) to make TyrRS_44CP1.

```
M. jannaschii TyrRS     KYVYGSEFQLDKDYTLNVYRLALKTTLKR-----ARRSMELIAREDENPKVAEVIYPIMQVN
TyrRS_44CP1             KYVYGSEFQWFGNMNVLTFLRDIGKHFSVNQMIDKEAVKQRLNREDQGISFTEVIYPIMQVN
```

Figure 3.2: Alignment of amino acid sequences between *M. jannaschii* TyrRS and TyrRS_44CP1. Amino acids 110-148 were removed from the *M. jannaschii* TyrRS and replaced with amino acids 129-172 from *E. coli* TyrRS to create TyrRS_44CP1. A TyrRS's CP1 domain directly interacts with the tyrosyl-tRNA's acceptor stem region.

If The CP1 substitution affected the *M. jannaschii* TyrRS in the same way it affected the human TyrRS then TyrRS_44CP1 should be able to recognize and charge a G1:C72 containing tyrosyl-tRNA. As detailed in Chapter 2, a plasmid (6x_1bp-$tRNA_{CUA}$) was created to express an orthogonal amber codon suppressing tRNA that was derived from *M. jannaschii* tyrosyl-tRNA. If the tRNA specificity of the TyrRS_44CP1 was changed as a result of the CP1 swap, then it should be able to recognize the G1:C72 containing 1bp-$tRNA_{CUA}$. We used a functional suppression assay to assess the activity of the TyrRS_44CP1 in HEK293T cells. An amber stop codon was first introduced into the 68^{th} position of a bacteriophage T4 (foldon) gene and cloned into pCDNA3.1 (Invitrogen, Carlsbad, CA). The gene also contained a sequence for a C-terminus V5 peptide that will only be translated if the amber stop codon is suppressed. 1bp-$tRNA_{CUA}$ is a functional suppressor tRNA only if it is charged with an amino acid. This can only occur if TyrRS_44CP1 recognizes it since we have previously shown that 1bp-$tRNA_{CUA}$ is not aminoacylated by endogenous HEK293T aaRS.

HEK293T cells were transfected with all combinations of three plasmids (TyrRS_44CP1, 6x_1bp-$tRNA_{CUA}$, foldon_68TAG). Cells were harvested 48 hours later and full length foldon protein was purified and detected using anti-V5 antibodies (**Figure 3.3**). Since the gene for the V5 tag is downstream of the amber stop codon, it will only be expressed if the amber stop codon is suppressed by a charged 1bp-$tRNA_{CUA}$.

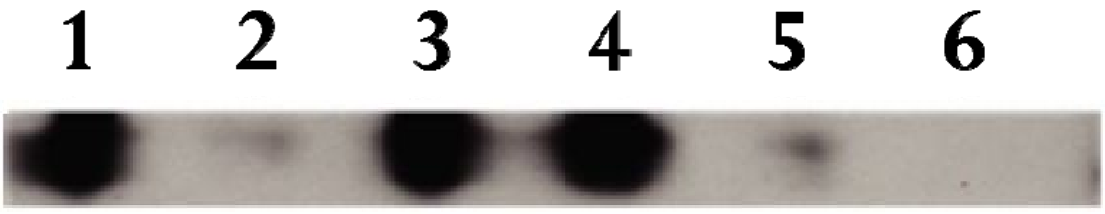


Figure 3.3: Detection of full length foldon protein using anti-V5 antibodies. HEK293T cells were transfected and harvested 48 hours later **(1)** Foldon_68TAG, 6x_wt-tRNA$_{CUA}$ **(2)** Foldon_68TAG, 6x_1bp-tRNA$_{CUA}$ **(3)** Foldon_68TAG, 6x_wt-tRNA$_{CUA}$, TyrRS_44CP1 **(4)** Foldon_68TAG, 6x_1bp-tRNA$_{CUA}$, TyrRS_44CP1 **(5)** Foldon_68TAG, TyrRS_44CP1 **(6)** Foldon_68TAG

Neither the presence of the TyrRS_44CP1 nor the 6x_1bp-tRNA$_{CUA}$ plasmid was able to cause the amber stop codon in the foldon gene to be suppressed (**Figure 3.3**). This demonstrates that each component is orthogonal in HEK293T cells. When both the TyrRS_44CP1 and 1bp-tRNA$_{CUA}$ were present in the cell, large amounts of full length foldon was detected. 1bp-tRNA$_{CUA}$ can only suppress the amber stop codon if it is charged with an amino acid. This shows that TyrRS_44CP1 can recognize and aminoacylate the G1:C72 containing 1bp-tRNA$_{CUA}$.

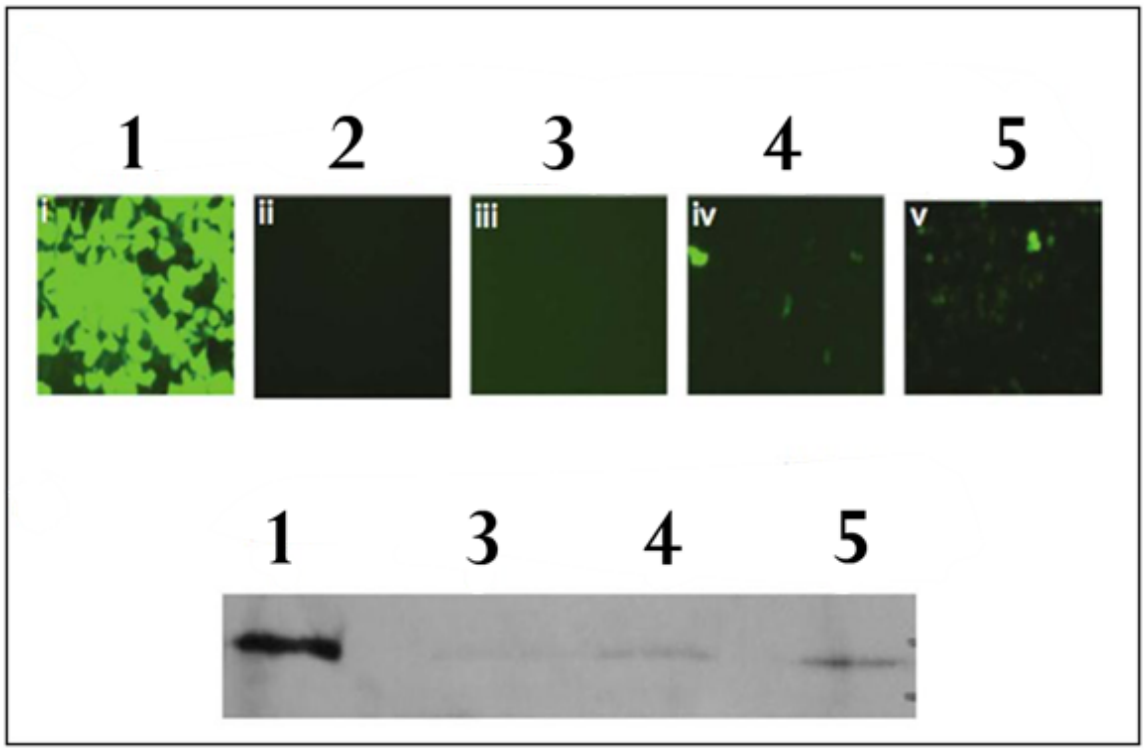


Figure 3.4: Detecting full length GFP using UV-light and anti-GFP antibodies. HEK293T cells were transfected and harvested 72 hours later **(1)** GFP **(2)** GFP_40TAG **(3)** GFP_40TAG, 6x_wt-tRNA$_{CUA}$,**(4)** GFP_40TAG, TyrRS_44CP1 **(5)** GFP_40TAG, 6x_wt-tRNA$_{CUA}$, TyrRS_44CP1

Green fluorescent protein (GFP) was also used as a reporter gene. Full length GFP can be easily visualized by UV-light or purified and detected with anti-GFP antibodies. An amber stop codon was introduced into the gene by changing the 40th codon from TAC (tyrosine) to TAG (amber stop codon). This completely eliminated the expression of full length GFP (**Figure 3.4**). HEK293T cells were also transfected with various combinations of the TyrRS_44CP1, 6x_1bp-tRNA$_{CUA}$, and GFP_40TAG plasmids in order to measure the suppression efficiency of the TyrRS_44CP1/6x_1bp-tRNA$_{CUA}$ pair. Combining the GFP_40TAG plasmid with either the 6x_1bp-tRNA$_{CUA}$ or TyrRS_44CP1 plasmid also did not restore the expression of full length GFP, although TyrRS_44CP1

and GFP40TAG did result in some increase. This again demonstrates that both components are orthogonal to host cell's endogenous tRNAs and aaRS. Co-transfecting HEK293T cells all three plasmids (6x_1bp-tRNA$_{CUA}$, TyrRS_44CP1, and GFP_40TAG) was the only case where a larger increase of full length GFP was detected. Analysis using fluorescence activated cell sorting (FACS) was also used to measure the suppression efficiency of the pair (restored ~1% of full length GFP expression).

Our results suggest that the *E. coli* TyrRS CP1 domain can be substituted into *M. jannaschii* TyrRS. This results in a functional TyrRS that can recognize and charge tyrosine to a G1:C72 containing tRNA (1bp-tRNA$_{CUA}$). Our approach can be applied to a nonnatural aaRS (evolved from *M. jannaschii* TyrRS) in order to manipulate it into charging our 1bp-tRNACUA with a nonnatural amino acid. However, the suppression efficiency of the pair is still low (1%) so subsequent sections will first focus on increasing the amounts of charged suppressor tRNA in the cell.

3.1.3 Materials and Methods

Cell Culturing and Transfections

All plasmids were amplified in *E. coli* TOP10 cells and isolated using Qiagen's Miniprep Kits. HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2. Transfections were performed according to product manual using FuGENE 6 (Roche, Indianapolis, IN). Full length GFP was excited at 460-500 nm and detected using a Nikon Eclipse TE2000-S microscope equipped with a FITC HyQ filter (Chroma).

Plasmid Construction

The 6x_1bp-tRNACUA, foldon_40TAG, and GFP_40TAG constructed in the same manner as previously described (**Chapter 2.1.3**). TyrRS_44CP1 was created by replacing the CP1 region of *M. jannaschii* TyrRS (amino acids 110-148) with the CP1 region of *E. coli* TyrRS (amino acids 129-172). The 5' end of the gene (containing sequences for amino acids 1-109) was amplified using the following primers:

5'-AAGGATCCACCATGGACGAATTTGAAATGAT-3'

5'-ACATTCATATTGCCGAACCACTGGAATTCACTTCCAT-3'

The 3' end (containing amino acids 149-306) was amplified by PCR using primers:

5'-AGGGGATTTCGTTCACTGAGGTTATCTATCCAATAATGCA-3'

5'-CCCGAATTCTAATCTCTTTCTAATTGGCT-3'

Finally, the *E. coli* CP1 region (amino acids 129-172) was amplified from a wild type *E. coli* TyrRS gene by PCR using the following primers:

5'-ATGGAAGTGAATTCCAGTGGTTCGGCAATATGAATGT-3'

5'-TGCATTATTGGATAGATAACTTCAGTGAACGAAATCCCCT-3'

All three PCR framers were combined, denatured for 15 minutes at 85 ºC and elongated with a klenow enzyme for 30 minutes at room temperature. The klenow product again amplified by PCR, purified, and digested with restriction enzymes HindIII and EcoRI. Finally, TyrRS_44CP1 was created by ligating the digested product into a pre-digested pEF6-V5-His6-TOPO (Invitrogen, Carlsbad, CA) using T4 DNA ligase (NEB, Ipswich, MA).

Harvest and Lysis of HEK293T Cells

72 hours after transfection, HEK293T cells were detached by physical scrapping and collected by centrifugation at 1000xg. A passive lysis buffer (Promega, Madison,

WI) supplemented with a Complete Mini-Protease inhibitor tablet (Roche, Indianapolis, IN) was used to lyse the collected cells on ice for 30 minutes. A soluble protein fraction was collected by centrifugation at 14,000 krpm for 20 minutes. The resulting solutible fraction was subjected to western blot analysis.

Western Blots

The soluble protein fractions were separated by 12% SDS-PAGE and transferred onto an Immobilon PVDF membrane (Millipore) in a transfer buffer (48 mM Tris, 39 mM glycine, 20% methanol (v/v), pH 9.2) using a BioRad Semi-Dry Blotter. The membrane was then probed for full length foldon expression using an anti-V5 antibody (Invitrogen, Carlsbad, CA) and a goat anti-mouse alkaline phosphatase conjugated secondary antibody (BioRad). The bands were visualized using a chemiluminscent reagent, PhosphaGlo AP (KPL) on a Kodak BioMax Light Film.

3.2 IMPROVING THE SUPPRESSION EFFICIENCY OF THE TYRRS_44CP1/1BP-TRNA$_{CUA}$ PAIR

3.2.1 Introduction

In the previous section we described how a CP1 substitution was used to manipulate *M. jannaschii* TyrRS into charging a G1:C72 containing tyrosyl-tRNA. The CP1 swapped TyrRS (TyrRS_44CP1) and an orthogonal amber codon suppressing tRNA (1bp-tRNA$_{CUA}$) was able to suppress an amber stop codon in HEK293T cells with an efficiency of ~1%. If this approach were applied to a nonnatural aaRS then we would only be able to express a very small amount of nonnatural amino acid incorporated protein. Therefore our next focus is on the optimization of the TyrRS_44CP1/1bp-tRNA$_{CUA}$ pair.

Suppression efficiency is directly related to the amount of charged (amino acid attached) suppressor tRNA present [4]. To achieve a level >1% we will first have to heighten the recognition and interaction between the TyrRS_44CP1 and 1bp-tRNA$_{CUA}$. A more active aaRS will result in an increase in the concentration of charged suppressor tRNA. We also sought to enhance the transcription and processing of the tRNA gene by HEK293T cells. Chapter 2 described how a human H1 promoter resulted in very high transcription of the 1bp-tRNA$_{CUA}$ gene. The effects of these improvements and more are investigated in this section.

3.2.2 Results and Discussion

Previous studies have suggested that the *M. jannaschii* TyrRS has very minimal interactions with the anticodon region of its cognate tRNA [5]. However, a recently published crystal structure of the *M. jannaschii* TyrRS complexed with its cognate tyrosyl-tRNA shows that some residues on the TyrRS do interact with the anticodon

region [6]. They also demonstrated that the recognition of a suppressor tRNA derived from *M. jannaschii* tyrosyl-tRNA could be improved by a factor of eight, *in vitro*, if the *M. jannaschii* TyrRS was mutated in one single residue. Thus, we generated a new mutant TyrRS gene derived from *M. jannaschii* TyrRS by first applying the same 44 amino acid transplantation as described in the previous section. The 286^{th} amino acid was mutated from aspartic acid to arginine as this was shown to increase the recognition of the CUA sequence on the suppressor tRNA's anticodon region [6]. This mutant gene was then cloned into pEF6-V5-His6-TOPO (Invitrogen, Carlsbad, CA) to make TyrRS_44CP1_R. An ochre stop codon (TAA) was inserted upstream of the V5 and His_6 tags sequences to prevent them from being translated (**Figure 3.5**).

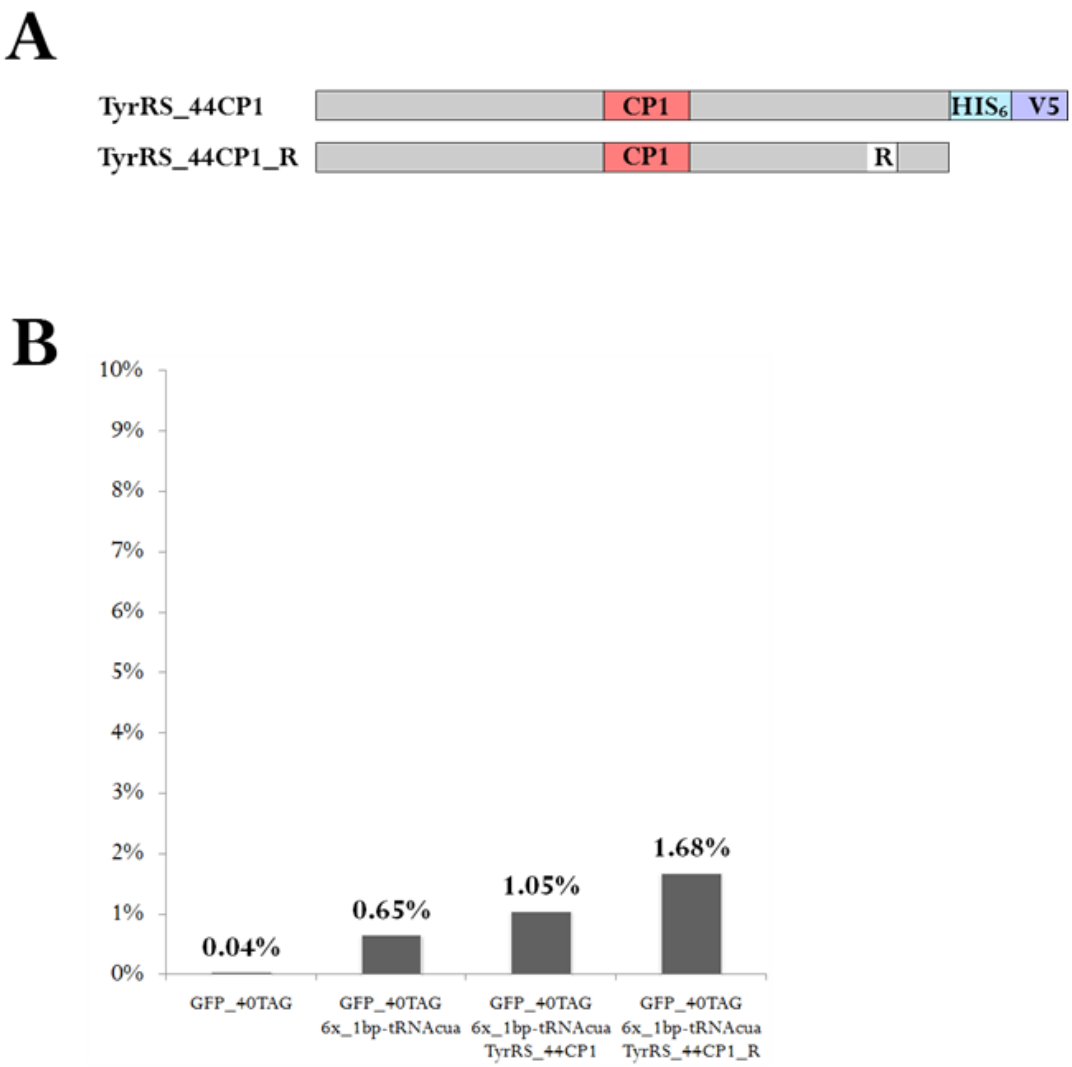


Figure 3.5: FACS was used to measure and quantify the percentage of positively fluorescent cells. TyrRS_44CP1_R resulted in a more than threefold increase over background when TyrRS_44CP1 only resulted in a two-fold increase. Arg286 enhances the recognition of the suppressor tRNA.

Once again a functional assay was developed to measure the suppression efficiency of the pair. An amber stop codon was introduced into a GFP gene in the 40th position to make GFP_40TAG. The tRNA plasmid (6x_1bp-tRNA$_{CUA}$) is described in Chapter 2. HEK293T cells were transfected with three plasmids: GFP_40TAG, 6x_1bp-

$tRNA_{CUA}$, and TyrRS_44CP1_ or TyrRS_44CP1_R. Cells were harvested 72 hours later and the total amount of full length GFP was measured using fluorescence activated cell sorting (FACS) (**Figure 3.5**). Some background, due to endogenous charging of the 1bp-$tRNA_{CUA}$, was observed but when TyrRS_44CP1 was added amount of full length GFP nearly doubled. When the C-terminus tags were removed and Arg286 introduced, TyrRS_44CP1_R was able to recognize 1bp-$tRNA_{CUA}$ almost 50% better. This evidence suggests that the Arg286 does enhance interactions between the aaRS and tRNA. In addition, the tags are unnecessary and probably only had a negative impact the TyrRS's activity. We also evaluated the effects of adding nuclear localization sequences to both the N-terminus and C-terminus ends of the CP1 swapped TyrRS. Sequence alignment of all twenty yeast aaRS has shown that fifteen of them contain classical NLS sequences [7]. It is also suggested that aminoacylation of the tRNA, inside the nucleus, is required for its export [8]. However, addition these NLS peptides resulted in non-functional enzymes (data not shown).

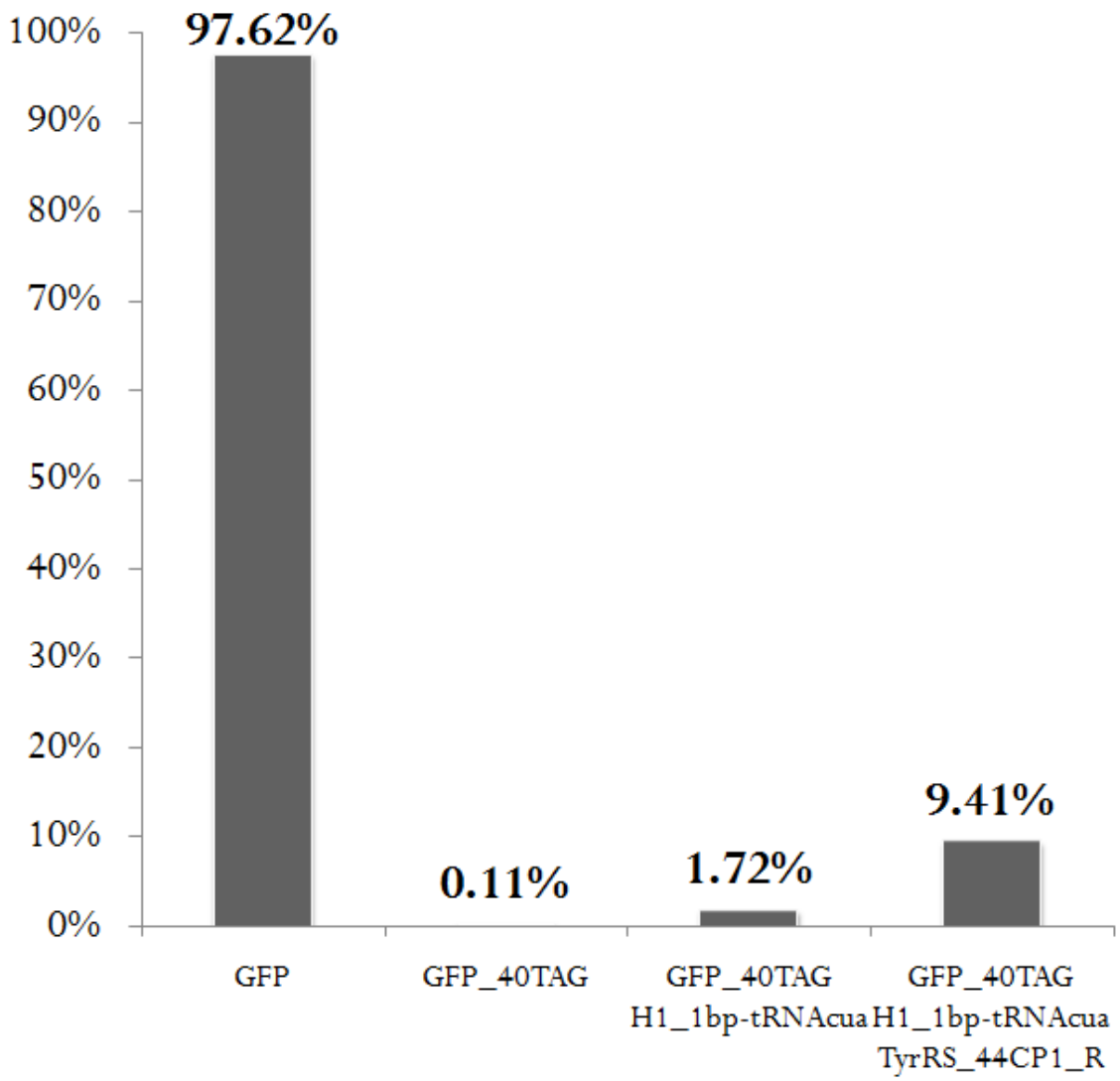


Figure 3.6: FACS was used to quantify the amount of fluorescent cells. Introducing an amber stop codon dramatically reduced the expression of full length, functional GFP. Addition of the 1bp-tRNA$_{CUA}$ resulted in a increase from charging by endogenous aaRS. When both components of the pair are present, a suppression efficiency of ~10% is observed.

Finally, we used the human H1 promoter (**Chapter 2**) to improve the transcription and processing of the 1bp-tRNA$_{CUA}$ gene. FACS was used to quantify the results of the functional assay (**Figure 3.6**). The amber stop codon prevented translation

of full length GFP and almost no fluorescent cells were observed. The 1bp-$tRNA_{CUA}$, while mostly orthogonal, is still recognized to some degree by endogenous aaRS in the HEK293T. The pair described in the previous section was able to achieve a suppression efficiency of only around 1%. Although our new pair greatly increased the background, by almost 250%, we also saw a 1000% increase in suppression efficiency. Chapter 4 will detail our efforts to genetically encode some nonnatural amino acids. High suppression efficiency will ensure that when we apply our CP1 approach to nonnatural aaRS sufficient amounts of incorporated protein can be produced.

3.2.3 Materials and Methods

Cell Culturing and Transfection of HEK293T Cells

Plasmids were amplified by transforming electrically component *E. coli* TOP10 cells, grown in LB, and isolated using QIAprep Spin Miniprep Kit (Qiagen, Madison, WI). HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2. Transfections complexes were formed using FuGENE HD (Roche, Indianapolis, IN). Plasmids were diluted in OPTI-MEM (Invitrogen, Carlsbad, CA) and FuGENE HD was added at a 2:1 (µL FuGENEHD: µg DNA). After 15 minutes, this was added to HEK293T cells (confluency 70-80%) 6 well flasks each containing 2 mL of media.

Plasmid Construction

GFP_40TAG, TyrRS_44CP1, 6x_1bp-tRNA$_{CUA}$, and H1_1bp-tRNA$_{CUA}$ plasmids were created as described in previous chapters. To make TyrRS_44CP1_R, the mutant TyrRS gene was amplified by PCR from TyrRS_44CP1 plasmid using the primers:

5'-GCGGATCCGCCACCATGGACGAATTTGAAATGATAAAGAGAA-'3

5'-GCTCTAGAGCTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTATA

AGTTCTTCAGCTACAGCATTTTTTAACCTCATTGGATGCAATTCCTT-'3

The PCR product contains a Kozak initiation sequence as well as the Arg286 mutation. Furthermore, an ochre stop codon (TAA) was added to prevent translation of the His_6 and V5 tags. After digesting with restriction enzymes BamHI and XbaI (New England Biolabs, Ipswich, MA) the product was ligated with a DNA T4 ligase into a predigested pEF6-V5-His6-TOPO (Invitrogen, Carlsbad, CA) to create TyrRS_44CP1_R.

SV40 NLS sequences (PKKKRKV) were added to either the N-terminus or C-terminus of the TyrRS_44CP1 gene by PCR using the following primers:

5'-CGGGATCCATGCCAAAAAAAAAAAGAAAAGTTGACGAATTTGAA

ATGATAAAG-3'

5'-GCTCTAGAGCTTAAACTTTTCTTTTTTTTTTTGGTAATCTCTTTCTAA

TTGGCTC-3'

PCR products were cleaned, digested, and ligated into pEF6-V5-His6-TOPO (Invitrogen, Carlsbad, CA).

FACS Analysis

A BD FACSCalibur flow cytometer was used to gate approximately 10,000 cells based on forward and side scatter. Each cell was then excited with 488 nm light and its resulting emission detected with an FL-1 filter (515-545 nm). Positive fluorescent

populations were gated based on negative controls or positive controls and analyzed using the computer software Cyflogic v.1.2.1 (CyFlo, Ltd., Turku, Finland).

3.3 INVESTIGATING OTHER CP1 SUBSTITUTIONS OF *M. JANNASCHII* TYRRS

3.3.1 Introduction

In the first two sections of this chapter, we showed how a 44 amino acid segment from *E. coli* TyrRS's CP1 domain can be used to change the tRNA specificity of *M. jannaschii* TyrRS. This CP1 substitution affected the orthogonality of the *M. jannaschii* TyrRS by changing its recognition of the tRNA's acceptor stem region from C1:G72 to G1:C72. Our orthogonal pair, TyrRS_44CP1/1bp-$tRNA_{CUA}$, was able to suppress an amber stop codon in HEK293T cells with an efficiency of ~10%. Before we apply this methodology to a nonnatural aaRS we wanted to see if we could identify more efficient peptide transplantations.

Wakasugi aligned amino acid sequence of *E. coli* TyrRS with human TyrRS in order to identify the proper CP1 transplantation size [3]. To choose our *M. jannaschii* TyrRS deletions we did a structural alignment of the crystal structures between human TyrRS and *M. jannaschii* TyrRS. In this section, we investigate three differ types of CP1 transplantation designs. Six are based off the structural alignment between *E. coli* TyrRS and *M. jannaschii* TyrRS and three based off of their amino acid sequence alignment. We also designed two different CP1 transplantations using sequences from *Thermus thermophilus* TyrRS. Each substitution is evaluated using a suppression assay for its recognition of 1bp-$tRNA_{CUA}$. This will allow us to identify an optimal CP1 substitution that can be applied to a nonnatural aaRS (**Chapter 4**).

3.3.2 Results and Discussion

In total, we designed eight CP1 substitutions that took sequences from *E. coli* and *Thermus thermophilus* TyrRS. TyrRS_44CP1_R was renamed TyrRS_44CP1 since all

designs also contained the Arg286 mutation that was shown to enhance recognition of the suppressor tRNA. Published crystal structures of the *E. coli* and *M. jannaschii* TyrRS revealed that both enzymes use a Rossmann-fold structural motif in order to bind the ATP [6, 9]. A region designated as the CP1 domain divides the six beta strands of the Rossmann fold into two parts and is involved in both dimerization as well as interacting directly with the tRNA's acceptor stem sequences. Alignment of the crystal structures between the *E. coli* (pdb.org 1X8X) and *M. jannaschii* TyrRS (pdb.org 1J1U) were used to design six *E. coli* CP1 swapped mutants: TyrRS_36CP1, TyrRS_39CP1, TyrRS_39CP1_2, TyrRS_42CP1, TyrRS_44CP1, and TyrRS_fullCP1. Next we used sequence alignment (CLUSTAW, Bioworkbench, SDSC) between the amino acids of the two TyrRS to design another three CP1 transplanted mutants: TyrRS_29RED, TyrRS_34RED, and TyrRS_38RED. The N-terminus of the CP1 swap are similar to the mutants designed with structural homology, however we choose the C-terminus based on a RED pattern that is shared by both enzymes. Finally, the crystal structure of *T. thermophilus* TyrRS is published and available [10]. We used structural alignment between it and the *M. jannaschii* TyrRS to design two final CP1 substituted mutants: TyrRS_39tt and TyrRS_45tt.

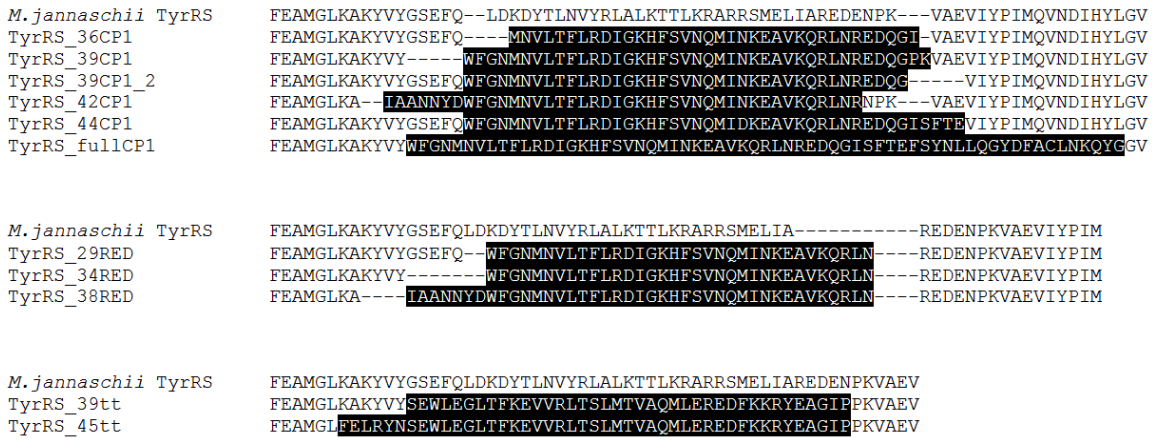


Figure 3.7: Aligning the amino acid sequences of the various CP1 transplanted mutants. Inserted CP1 domains are shown in black. Six used sequences from *E. coli* TyrRS while two were taken from *T. thermophilus* TyrRS. Each was tested for their ability to charge 1bp-tRNA$_{CUA}$ in HEK293T cells. (Bioworkbench, SDSC)

A suppression assay was used to evaluate the activity of each CP1 substituted *M. jannaschii* TyrRS for our suppressor tRNA (1bp-tRNA$_{CUA}$). An amber stop codon was introduced into the 40th position of a GFP gene to make the plasmid GFP_40TAG. In Chapter two we described our design and creation of H1_1bp-tRNA$_{CUA}$, which can express an orthogonal amber codon suppressing tRNA in mammalian cells. HEK293T cells were transfected with three plasmids: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, and a plasmid containing the CP1 swapped TyrRS gene. If the 1bp-tRNA$_{CUA}$ can be charged by the mutant TyrRS, then a functional suppression tRNA will be present in the cell. As a result, the amber stop codon in the GFP_40TAG gene can be suppressed and full length GFP expressed. A non-charged 1bp-tRNA$_{CUA}$ would not be able to suppress the amber stop codon, resulting in the translation of a truncated and non-functional form of GFP.

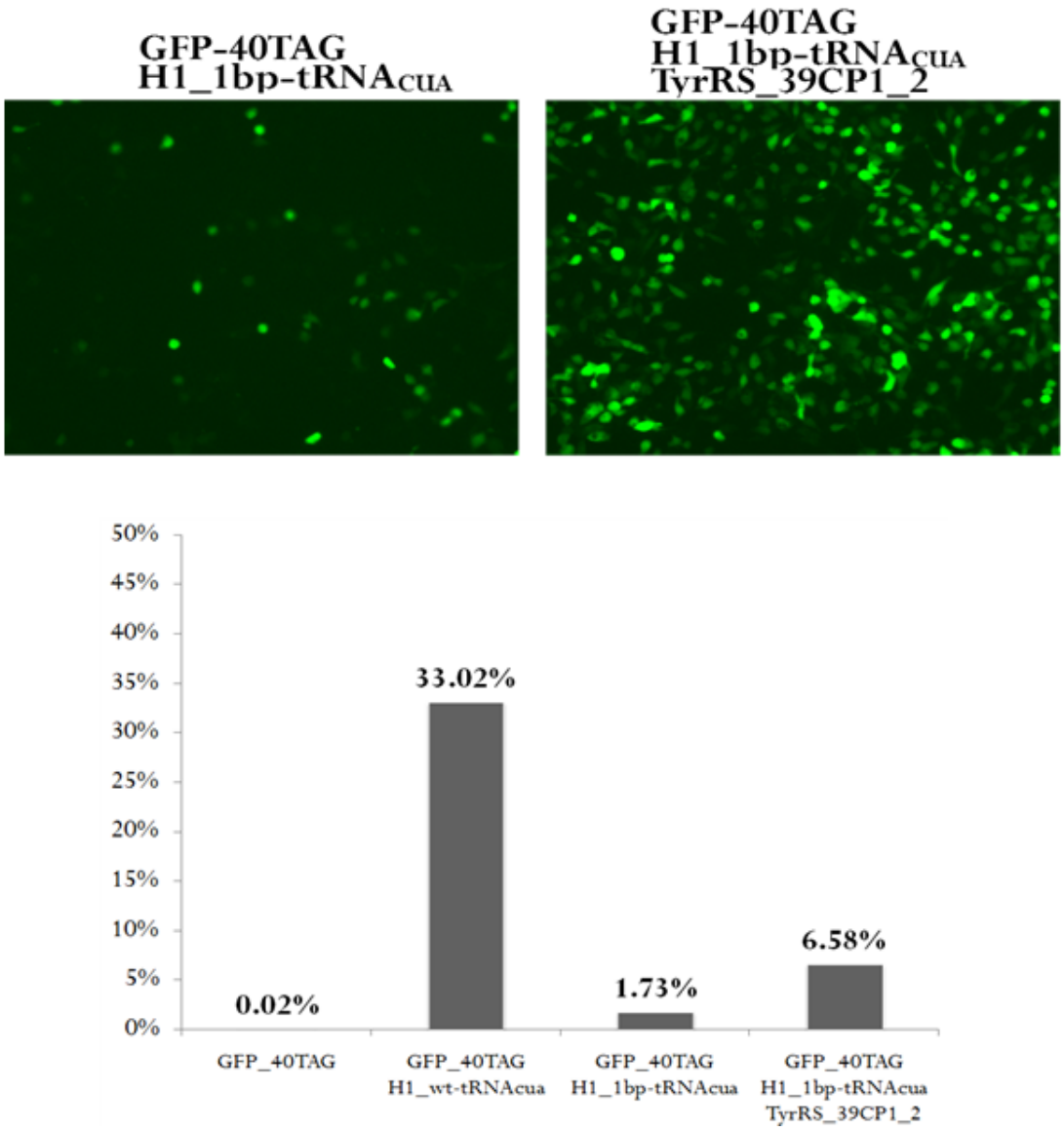


Figure 3.8: Using a functional assay to test mutant TyrRS's recognition of 1bp-$tRNA_{CUA}$. FACS was used to measure the percentage of green fluorescing cells 72 hours after transfection. The TyrRS_39CP1_2/1bp-$tRNA_{CUA}$ pair was able to restore about >6% of full length GFP expression. This is roughly a 380% increase over the 1bp-$tRNA_{CUA}$ background. All eight CP1 substituted TyrRS variations were tested in a similar manner.

Figure 3.8 shows the results of our testing of the TyrRS_39CP1_2/1bp-$tRNA_{CUA}$ pair. Almost no full length GFP was observed when cells were only transfected with

GFP_40TAG. Addition of the 1bp-tRNA$_{CUA}$ resulted in some increase which is consistent with previous results (**Chapter 2**). However, when both elements of the pair were present the amount of full length GFP increased by about 380% (ratio 3.8) compared to just 1bp-tRNA$_{CUA}$ alone. All of the eight CP1 mutants were evaluated using a similar functional assay. *E. coli* TyrRS used as a positive control since the *prokaryotic* TyrRS recognizes a G1:C72 sequence on its corresponding tyrosyl-tRNA and is able to charge 1bp-tRNA$_{CUA}$ with high efficiency (**Chapter 2**). Pairing it up with 1bp-tRNA$_{CUA}$ resulted in a 13.5 factor increase over 1bp-tRNA$_{CUA}$ background (**Figure 3.9**)

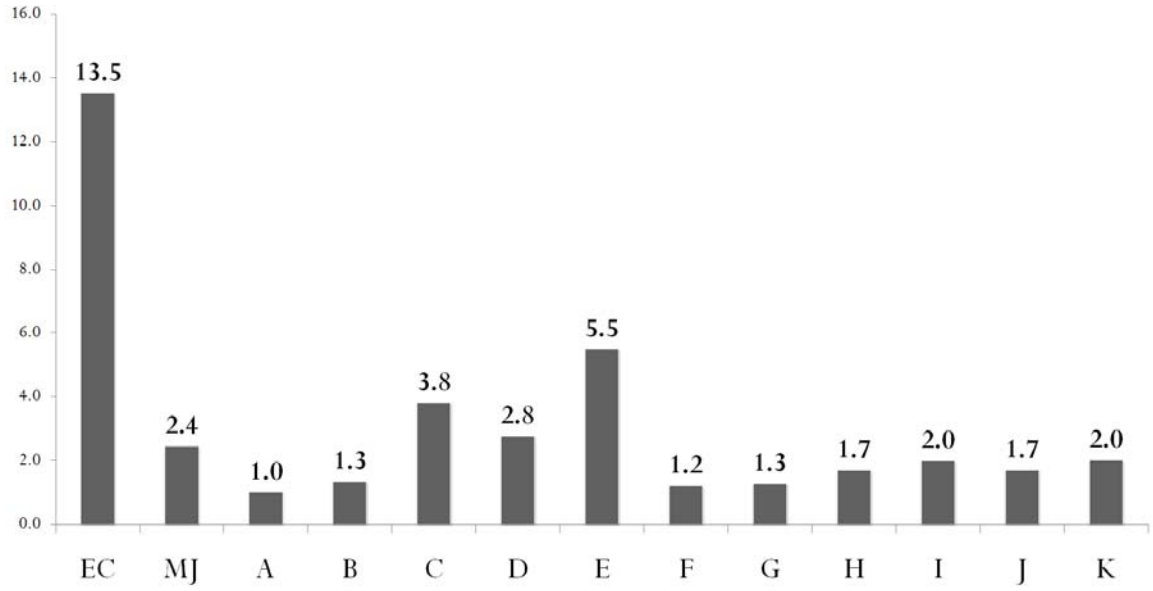


Figure 3.9: Ratio of suppression efficiency to tRNA background. A functional assay tested the recognition of 1bp-tRNA$_{CUA}$ by a CP1 substituted TyrRS. 1bp-tRNA$_{CUA}$ background is always ~1.7% of gated cells. *E. coli* TyrRS-**(EC)** *M. jannaschii* TyrRS-**(MJ)** TyrRS_36CP1-**(A)** TyrRS_39CP1-**(B)** TyrRS_39CP1_2-**(C)** TyrRS_42CP1-**(D)** TyrRS_44CP1-**(E)** TyrRS_fullCP1-**(F)** TyrRS_29RED-**(G)** TyrRS_34RED-**(H)**, TyrRS_38RED-**(I)** TyrRS_39tt-**(J)** TyrRS_45tt-**(K)**

Wild type *M. jannaschii* TyrRS was also used as a negative control. Not surprising, it was still able to recognize the G1:C72 containing 1bp-$tRNA_{CUA}$ and resulted in a twice the number of fluorescent cells. Some variations were barely functional and did not result in any increase: TyrRS_36CP1 and TyrRS_fullCP1. We were surprised that a full length CP1 swap (TyrRS_fullCP1) performed so poorly since a full CP1 swapped human TyrRS was functional [3]. Poor suppression efficiency could be caused by the enzyme misfolding. If the CP1 substitution disrupted the TyrRS's active site then a nonfunctional enzyme would result. None of the *T. thermophilus* CP1 transplantations recognized 1bp-$tRNA_{CUA}$ very well either. Our understanding of the *E. coli* TyrRS's CP1 domain is more detailed and probably yielded more rationally designed substitutions. The best CP1 substitution was still the TyrRS_44CP1 identified in the previous section although TyrRS_39AA_2 also resulted in an almost four times increase in fluorescing cells. This concludes our work on changing the orthogonality of *M. jannaschii* TyrRS. The most effective CP1 substitutions can be applied to a nonnatural aaRS to change its orthogonality. Manipulating it nonnatural aaRS into charging 1bp-$tRNA_{CUA}$, without negatively affecting its amino acid recognition can result in functional orthogonal nonnatural aaRS. Chapter 4 will focus on our efforts to expand the genetic code in mammalian cells by applying this technique to site-specifically incorporate a nonnatural amino acid.

3.3.3 Materials and Methods

Cell Culturing and Transfections

All plasmids were amplified in *E. coli* TOP10 cells and isolated using Qiagen's Miniprep Kits. HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's

Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2. Transfections were performed according to product manual using FuGENE HD (Roche, Indianapolis, IN). Full length GFP was excited at 460-500 nm and detected using a Nikon Eclipse TE2000-S microscope equipped with a FITC HyQ filter (Chroma).

Plasmid Construction

GFP_40TAG, *E. coli* TyrRS, TyrRS_44CP1_R, 6x_1bp-$tRNA_{CUA}$, and H1_1bp-$tRNA_{CUA}$ plasmids were created as described in previous chapters. All CP1 transplanted mutants were created by PCR in a process similar to the one described in the previous section. The chimeric gene was divided into three segments, each one amplified by PCR. Finally, a joint gene was also amplified by PCR, digested, and ligated into pEF6-V5-His6-TOPO (Invitrogen, Carlsbad, CA). Primers are listed for each of the segments based on the size and sequence of the CP1 swap.

TyrRS_36CP1:

5’-AAGGATCCACCATGGACGAATTTGAAATGAT-3’

5’-CGCAGGAAGGTCAGCACATTCATCTGGAATTCACTTCCATAAA CAT-3’

5’-ATGTTTATGGAAGTGAATTCCAGATGAATGTGCTGACCTTCCT GCG-3’

5’- TGGATAGATAACTTCAGCAACAATCCCCTGATCTTCACGGTT GAGA-3’

5’-TCTCAACCGTGAAGATCAGGGGATTGTTGCTGAAGTTATC TATCCA-3’

5’-TGCATTATTGGATAGATAACTTCAGTGAACGAAATCCCCT-3’

TyrRS_39CP1:

5’-AAGGATCCACCATGGACGAATTTGAAATGAT-3’

5’-ACATTCATATTGCCGAACCAATAAACATATTTTGCCTT-3’

5’-AGGCAAAATATGTTTATTGGTTCGGCAATATGAATGT-3’

5’-ATAACTTCAGCAACCTTTGGCCCCTGATCTTCACGGTTGA-3’

5’-TCAACCGTGAAGATCAGGGGCCAAAGGTTGCTGAAGTTAT-3’

5’-TGCATTATTGGATAGATAACTTCAGTGAACGAAATCCCCT-3’

TyrRS_39CP1_2:

5’-GCGGATCCGCCACCATGGACGAATTTGAAATGATAAAGAGAA-‘3

5’-ACATTCATATTGCCGAACCAATAAACATATTTTGCCTT-‘3

5’-AAGGCAAAATATGTTTATTGGTTCGGCAATATGAATGT-‘3

5’-ATAACTTCAGCAACCTTTGGCCCCTGATCTTCACGGTTGA-‘3

5’-TCAACCGTGAAGATCAGGGGCCAAAGGTTGCTGAAGTTAT-‘3

5’-GCTCTAGAGCTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTATA

AGTTCTTCAGCTACAGCATTTTTTAACCTCATTGGATGCAATTCCTT-‘3

TyrRS_42CP1:

5’-GCGGATCCGCCACCATGGACGAATTTGAAATGATAAAGAGAA-‘3

5’-TCATAGTTGTTCGCCGCGATTGCCTTTAACCCCATTGCTT-‘3

5’-AAGCAATGGGGTTAAAGGCAATCGCGGCGAACAACTATGA-‘3

5’-TTCAGCAACCTTTGGATTACGGTTGAGACGCTGCTTAACC-‘3

5’-GGTTAAGCAGCGTCTCAACCGTAATCCAAAGGTTGCTGAA-‘3

5’-GCTCTAGAGCTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTATA

AGTTCTTCAGCTACAGCATTTTTTAACCTCATTGGATGCAATTCCTT-‘3

TyrRS_fullCP1:

5’-AAGGATCCACCATGGACGAATTTGAAATGAT-3’

5’-ACATTCATATTGCCGAACCAATAAACATATTTTGCCTT-3’

5’-AGGCAAAATATGTTTATTGGTTCGGCAATATGAATGT-3’

5’-CCAACTGCAACATCAACGCCACCGTACTGTTTGTTCAGAC-3’

5’-GTCTGAACAAACAGTACGGTGGCGTTGATGTTGCAGTTGG-3’

5’-TGCATTATTGGATAGATAACTTCAGTGAACGAAATCCCCT-3’

TyrRS_29RED:

5’-GCGGATCCGCCACCATGGACGAATTTGAAATGATAAAGAGAA-‘3

5’-ACATTCATATTGCCGAACCAATAAACATATTTTGCCTT-‘3

5’-AAGGCAAAATATGTTTATTGGTTCGGCAATATGAATGT-‘3

5’-TTGGATTTTCATCCTCTCTGTTGAGACGCTGCTTAACCGC-‘3

5’-GCGGTTAAGCAGCGTCTCAACAGAGAGGATGAAAATCCAA-‘3

5’-GCTCTAGAGCTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTATA

AGTTCTTCAGCTACAGCATTTTTTAACCTCATTGGATGCAATTCCTT-‘3

TyrRS_34RED:

5’-AAGGATCCACCATGGACGAATTTGAAATGAT-3’

5’-ACATTCATATTGCCGAACCAATAAACATATTTTGCCTT-3’

5’-AGGCAAAATATGTTTATTGGTTCGGCAATATGAATGT-3’

5’-TTGGATTTTCATCCTCTCTGTTGAGACGCTGCTTAACCGC-3’

5’-GCGGTTAAGCAGCGTCTCAACAGAGAGGATGAAAATCCAA-3’

5’-TGCATTATTGGATAGATAACTTCAGTGAACGAAATCCCCT-3’

TyrRS_38RED:

5’-GCGGATCCGCCACCATGGACGAATTTGAAATGATAAAGAGAA-‘3

5’-TCATAGTTGTTCGCCGCGATTGCCTTTAACCCCATTGCTT-‘3

5’-AAGCAATGGGGTTAAAGGCAATCGCGGCGAACAACTATGA-‘3

5’-TTGGATTTTCATCCTCTCTGTTGAGACGCTGCTTAACCGC-‘3

5’-GCGGTTAAGCAGCGTCTCAACAGAGAGGATGAAAATCCAA-‘3

5’-GCTCTAGAGCTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTATA

AGTTCTTCAGCTACAGCATTTTTTAACCTCATTGGATGCAATTCCTT-‘3

TyrRS_39tt:

5’-AAGGATCCACCATGGACGAATTTGAAATGAT-3’

5’-GTGAGGCCCTCCAGCCACTCGGAATAAACATATTTTGCCTTTAAC-3’

5’-GTTAAAGGCAAAATATGTTTATTCCGAGTGGCTGGAGGGC

CTCAC-3’

5’-ATAACTTCAGCAACCTTTGGGGGAATCCCCGCCTCGTACCGCTTC-3’

5’-GAAGCGGTACGAGGCGGGGATTCCCCCAAAGGTTGCTGAAG

TTAT-3’

5’-TGCATTATTGGATAGATAACTTCAGTGAACGAAATCCCCT-3’

TyrRS_45tt:

5’-AAGGATCCACCATGGACGAATTTGAAATGAT-3’

5’-CTCGGAGTTGTAGCGGAGCTCAAATAACCCCATTGCTTCAAA

AAC-3’

5’-GTTTTTGAAGCAATGGGGTTATTTGAGCTCCGCTACAACTCCGAG-3’

5’-ATAACTTCAGCAACCTTTGGGGGAATCCCCGCCTCGTACCGCTTC-3’

5’-GAAGCGGTACGAGGCGGGGATTCCCCCAAAGGTTGCTGAAG

TTAT-3’

5’-TGCATTATTGGATAGATAACTTCAGTGAACGAAATCCCCT-3’

FACS Analysis

A BD FACSCalibur flow cytometer was used to gate approximately 10,000 cells based on forward and side scatter. Each cell was then excited with 488 nm light and its

resulting emission detected with an FL-1 filter (515-545 nm). Positive fluorescent populations were gated based on negative controls or positive controls and analyzed using the computer software Cyflogic v.1.2.1 (CyFlo, Ltd., Turku, Finland).

3.4 REFERENCES

1. Liu, D.R., et al., *Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo.* Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10092-7.
2. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.
3. Wakasugi, K., et al., *Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant.* EMBO J, 1998. **17**(1): p. 297-305.
4. Goodman, H.M., et al., *Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA.* Nature, 1968. **217**(5133): p. 1019-24.
5. Steer, B.A. and P. Schimmel, *Major anticodon-binding region missing from an archaebacterial tRNA synthetase.* J Biol Chem, 1999. **274**(50): p. 35601-6.
6. Kobayashi, T., et al., *Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion.* Nat Struct Biol, 2003. **10**(6): p. 425-32.
7. Schimmel, P. and C.C. Wang, *Getting tRNA synthetases into the nucleus.* Trends Biochem Sci, 1999. **24**(4): p. 127-8.
8. Lund, E. and J.E. Dahlberg, *Proofreading and aminoacylation of tRNAs before export from the nucleus.* Science, 1998. **282**(5396): p. 2082-5.

9. Kobayashi, T., et al., *Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase.* J Mol Biol, 2005. **346**(1): p. 105-17.

10. Yaremchuk, A., et al., *Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition.* EMBO J, 2002. **21**(14): p. 3829-40.

Chapter Four

Genetically Encoding a Nonnatural Amino Acid

4.1 Site-specific incorporation of 3,4-dihydroxy-L-phenylalanine

4.1.1 Introduction

Chapter 1 reviewed several techniques that are currently employed to incorporate nonnatural amino acids into a protein. Schultz's group used pairs that consisted of a suppressor tRNA and a corresponding nonnatural aminoacyl-tRNA synthetase (aaRS). When introduced into an orthogonal host, these pairs can be used to site-specifically incorporate a nonnatural amino acid in response to the amber stop codon (TAG). Dozens of nonnatural aaRS have been evolved from *Methanococcus jannaschii* tyrosyl-tRNA synthetase (TyrRS). These pairs have allowed many nonnatural amino acids (>50) to be successful incorporated into a protein in *prokaryotic* cells but only a few (<10) are available in mammalian systems [1]. Our project is focused on the development of more mammalian useable pairs. In the previous two chapters, we concentrated our efforts on altering the *M. jannaschii* tyrosyl-tRNA and TyrRS pair so that is orthogonal in mammalian cells. In this chapter we use our connective polypeptide 1 (CP1) substitution (**Chapter 3**) to change the orthogonality of several nonnatural aaRS derived from *M. jannaschii* TyrRS. If the CP1 substitution does not affect the nonnatural aaRS's amino acid specificity, then it should be able to charge our 1bp-$tRNA_{CUA}$ with a nonnatural amino acid.

3,4-dihydroxy-L-phenylalanine (L-DOPA) is a nonnatural amino acid analog of tyrosine (**Figure 4.1**) that can be used to capture transient protein-protein interactions. Our lab has previously incorporated this nonnatural amino acid in response to an amber

stop codon in a gene in *Escherichia coli* cells [2]. We used a *M. jannaschii* TyrRS derived nonnatural aaRS that was previously evolved by Alfonta which specifically recognizes L-DOPA. The mutant *M. jannaschii* TyrRS contained four amino acid changes in its active site (Glu26, Leu32, Ser67, Asn70, and Gln167) [3]. First we substitute the CP1 domain of this L-DOPA specific aaRS with the CP1 region from *E. coli* TyrRS. We then investigated the ability of this nonnatural aaRS to recognize and charge our 1bp-$tRNA_{CUA}$ with L-DOPA in HEK293T cells. If its active site is not negatively affected by the CP1 substitution, then it should still be able to recognize the nonnatural amino acid while discriminating against tyrosine. Introduction of this orthogonal pair can be used to incorporate L-DOPA into protein whenever an amber stop codon in present in a gene.

4.1.2 Results and Discussion

M. jannaschii TyrRS evolved L-DOPA aaRS, L-DOPA_RS, was first modified in its connective polypeptide 1 (CP1) region. Amino acids 110-148 were removed from L-DOPA_RS and replaced with amino acids 129-167 from *Escherichia coli* TyrRS to change its tRNA recognition from C1:G72 to G1:C72 (as discussed in **Chapter 3**). We believe that this will allow the L-DOPA_ 39CP1 to recognize our orthogonal amber codon suppressing tRNA, 1bp-$tRNA_{CUA}$, and charge it with L-DOPA. Finally, a reporter gene was constructed from green fluorescent protein (GFP) by introducing an amber stop codon in the 40^{th} position. After transfection HEK293T cells with all three plasmids, expression of full length GFP will only occur if the cell has an aminoacylated suppressor tRNA. 1bp-$tRNA_{CUA}$ is orthogonal in HEK293T cells (**Chapter 2**) so only L-DOPA_39CP1 can recognize and charge it with L-DOPA. We use UV-light to detect full

length GFP and used fluorescence activated cell sorting (FACS) to measure the population of fluorescent cells.

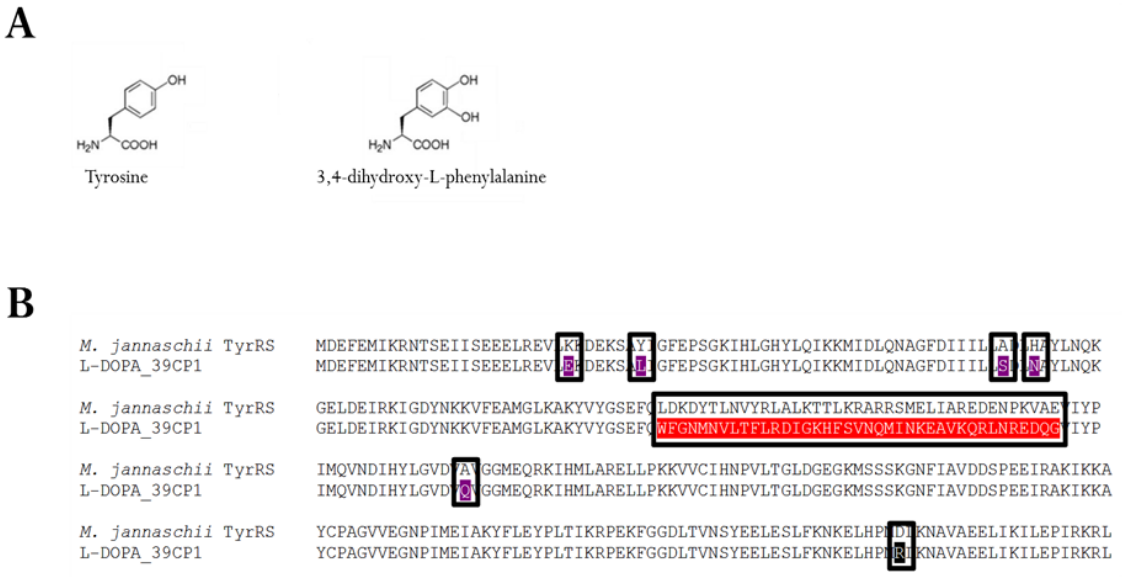


Figure 4.1: (A) Tyrosine and L-DOPA are only different in their side chain. The di-hydroxy functional group is not found on any of the twenty natural amino acids. **(B)** Primary sequence alignment of the L-DOPA_39CP1 and wild type *M. jannaschii* TyrRS. Glu26, Leu32, Ser67, Asn70, and Gln167 allow it to recognize L-DOPA. A CP1 substitution (red) changes its acceptor stem recognition from C1:G72 to G1:C72. Arg286 improves its recognition of the suppressor 1bp-tRNA$_{CUA}$'s anticodon region.

HEK293T cells were transfected with the GFP_40TAG, H1_1bp-tRNA$_{CUA}$, and L-DOPA_39CP1 plasmids. 96 hours later, the cells were harvested and full length GFP was detected using FACS (**Figure 4.2**). Since only charged suppressor tRNAs can be used by the cell to translate full length GFP, the percentage of fluorescent cells was used to measure the activity of the L-DOPA_39CP1 for the 1bp-tRNA$_{CUA}$. When cells were transfected with the H1_1bp-tRNA$_{CUA}$ plasmid only a small amount of suppression efficiency was observed. This indicates that the 1bp-tRNA$_{CUA}$ is mostly orthogonal,

although some recognition by endogenous TyrRS does exist. The addition of the L-DOPA_39CP1 actually resulted in a more than threefold increase in full length GFP expression even when L-DOPA was not present in the growth medium. Unfortunately, this shows that the L-DOPA_39CP1 synthetase is still able to recognize tyrosine or some other natural amino acid and charge it to the 1bp-$tRNA_{CUA}$. The CP1 substitution is affecting the amino acid specificity and reducing the fidelity of the L-DOPA_39CP1 for its nonnatural amino acid. When L-DOPA was supplied to the growth medium (5 mM) we actually saw a small decrease in the amount of full length GFP. When we superimposed the FACS data between when L-DOPA was present or absent we can see that that the percentage of lower fluorescing cells actually increased while higher fluorescing cells slightly decreased. We can speculate on some of the possible reasons. L-DOPA could be negatively affecting health of the cells. To dissolve L-DOPA into the media we had to reduce the pH of the media to about 7.0 compared to normal media pH which is 7.5. As a result, protein expression could be reduced as a result of an unhealthy cell. Another possibility is if L-DOPA incorporation into GFP is causing it to be degraded more rapidly. Since the FACS data was inconclusive, we decided to use redox-cycling staining assay to detect L-DOPA incorporated GFP.

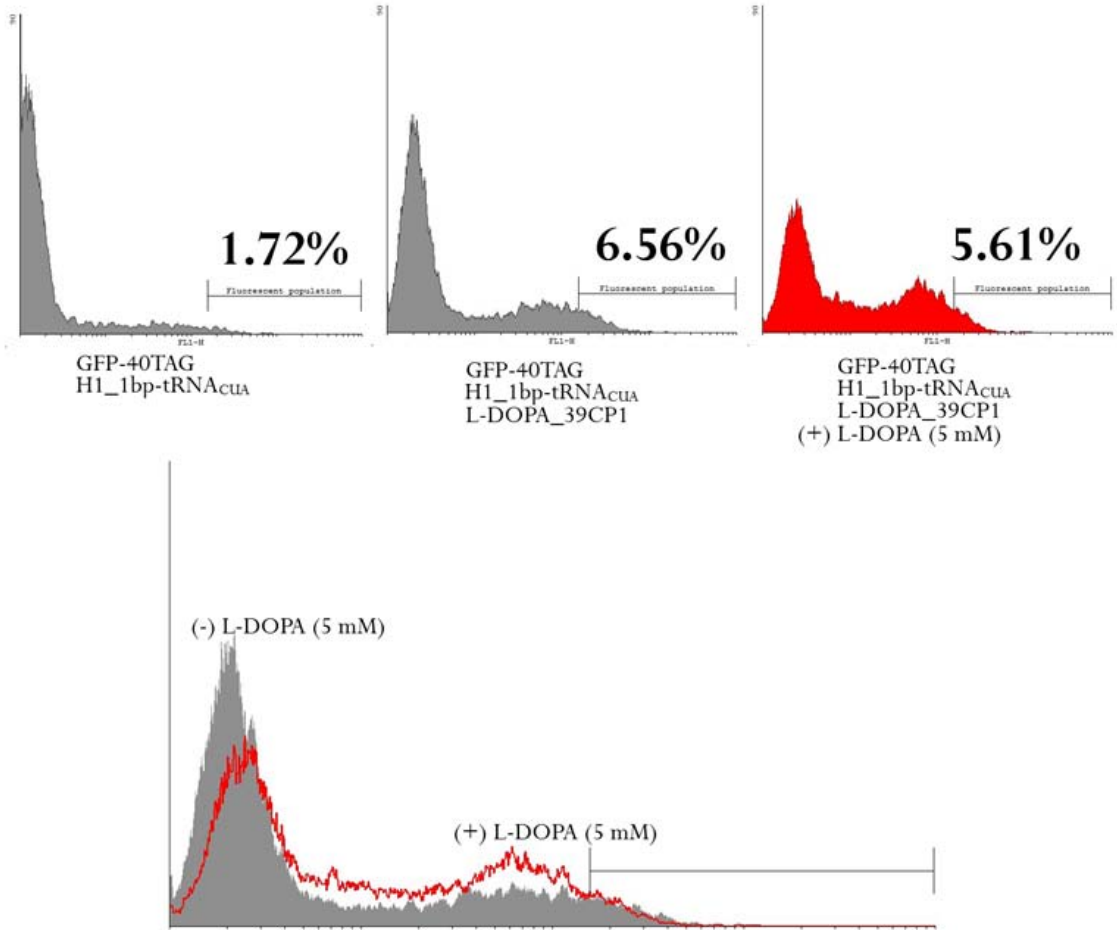


Figure 4.2: A functional assay was used to assess if L-DOPA_39CP1 could aminoacylate 1bp-tRNA$_{CUA}$ with L-DOPA in HEK293T cells. L-DOPA_39CP1 charged 1bp-tRNA$_{CUA}$ in the absence of L-DOPA. When L-DOPA (5 mM) was supplied to the growth media, the cells containing low amount of full length GFP increased while cells that had high amounts of GFP slightly decreased.

As before, HEK293T cells were transfected with the three plasmids: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, and L-DOPA_39CP1. L-DOPA was also supplied to the cells at 5 mM concentration. 96 hours after transfection, full length GFP was purified using Ni-NTA agarose beads (Qiagen, Madison, WI). The sample was then concentrated, resolved by SDS polyacrylamide gel electrophoresis (PAGE) transferred to a

nitrocellulose membrane, and full length GFP was detected using anti-GFP antibodies (**Figure 4.3**). A control lane of over expressed wild type full length GFP was used as a positive control.

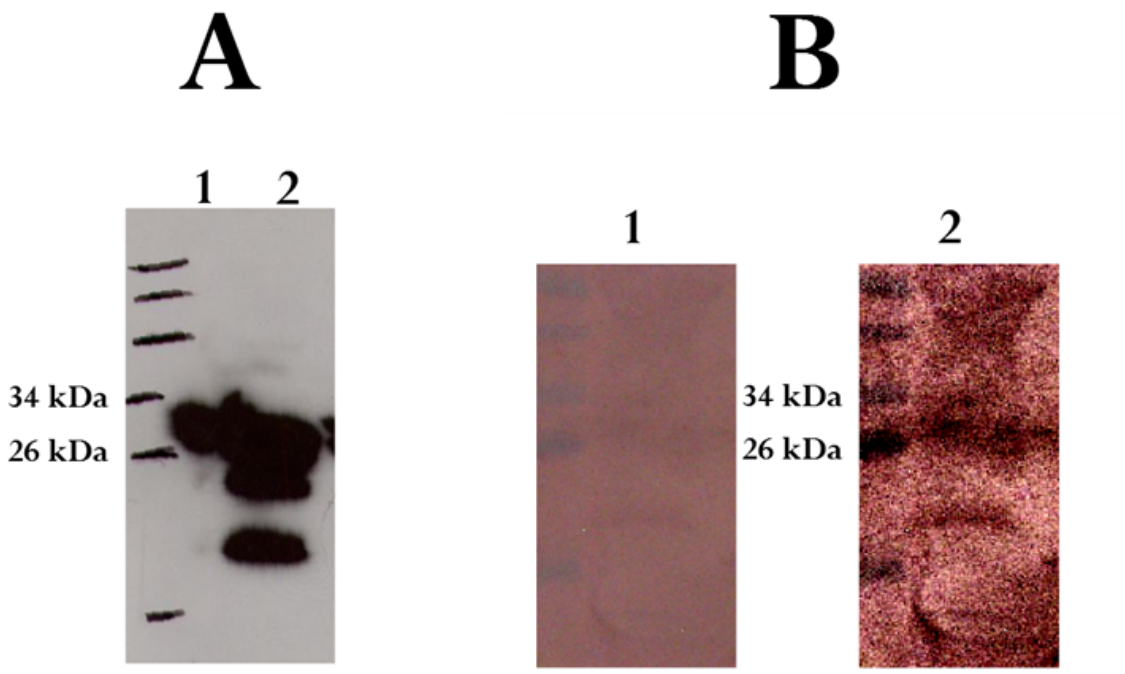


Figure 4.3: Detection of L-DOPA incorporated GFP using anti-GFP antibodies and redox-cycling staining. (A) Full length GFP was detected with western blot using anti-GFP antibodies. Lane 1 shows wild type GFP. Lane 2 shows the L-DOPA incorporated GFP. (B) Shows the results of NBT staining for L-DOPA. Low signals are enhanced in Lane 2 with Adobe Flash CS5.

L-DOPA incorporated protein was also detected using nitro blue tetrazolium chloride (NBT) in a redox-cycling staining method [2, 4]. A very faint band was detected at the molecular weight of GFP (28 kDa). (**Figure 4.3**) In addition, lower molecular weight products were also detected. The staining data corresponds to the western blot which seems to full length GFP is present in addition to some lower molecular weight

products. This in addition to the FACS data seems to suggest that L-DOPA is being incorporated into the GFP.

We have demonstrated that the previously evolved *M. jannaschii* derived L-DOPA pair can be used in mammalian cells with some modification. The CP1 transplantation allowed the nonnatural aaRS to recognize and charge 1bp-$tRNA_{CUA}$. When L-DOPA was supplied to the cells, we saw some level of nonnatural amino acid incorporation. Since the FACS data does not reveal the identity of the incorporated amino acid so we used a staining assay to detect the L-DOPA. Promising results from the both genetic and biochemical analysis demonstrated the validity of our technique. This approach can be applied to other *M. jannaschii* derived nonnatural synthetases for rapid expansion of the mammalian genetic code.

4.1.3 Materials and Methods

Cell Culturing and Transfection of HEK293T Cells

Plasmids were amplified by transforming electrically component *E. coli* TOP10 cells, grown in LB, and isolated using QIAprep Spin Miniprep Kit (Qiagen, Madison, WI). HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2. Transfections complexes were formed using FuGENE HD (Roche, Indianapolis, IN). Plasmids were diluted in OPTI-MEM (Invitrogen, Carlsbad, CA) and FuGENE HD was added at a 2:1 (µL FuGENE HD: µg DNA). After 15 minutes, this was added to HEK293T cells (confluency 40-60%) in T-75 culture flasks.

Plasmid Construction

The construction of the GFP_40TAG and H1-1bp-tRNA$_{CUA}$ plasmids is described in previous chapters. L-DOPA_39CP1 was generated by replacing amino acid 110-148 in the L-DOPA specific aaRS with amino acids 129-167 from *E. coli* TyrRS. The 5' segment (segment 1) of a previously evolved L-DOPA aaRS gene was amplified from the plasmid pAC-DHPhe-6TRN (Gift from Dr. Peter G. Schultz, The Scripps Research Institute, La Jolla, CA) with the following primers:

5'-GCGGATCCGCCACCATGGACGAATTTGAAATGATAAAGAGAA'3

5'-ACATTCATATTGCCGAACCAATAAACATATTTTGCCTT-'3

A sequence (segment 2) for *E. coli* TyrRS's CP1 domain was amplified by PCR using:

5'-AAGGCAAAATATGTTTATTGGTTCGGCAATATGAATGT-'3

5'-ATAACTTCAGCAACCTTTGGCCCCTGATCTTCACGGTTGA-'3

Finally, the 3' fragment (segment 3) of the L-DOPA aaRS was amplified from pAC-DHPhe-6TRN using:

5'-TCAACCGTGAAGATCAGGGGCCAAAGGTTGCTGAAGTTAT-'3

5'-GCTCTAGAGCTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTATA

AGTTCTTCAGCTACAGCATTTTTTAACCTCATTGGATGCAATTCCTT-'3

The segment 1 forward and segment 2 reverse primers were used to amplify a sequence containing segments 1 and 2. Finally, this segment 1_2 joined to segment 3 by PCR using segment 1 forward and segment 3 reverse primers. This gene also contained a Kozak initiation sequence and the Arg286 mutation. The segment 1_2_3 PCR product was digested with BamHI and XbaI (New England Biolabs, Ipswich, MA) and ligated using DNA T4 Ligase into a predigested pEF6-V5 (Invitrogen, Carlsbad, CA) to create L-DOPA_39CP1.

Genetic Incorporation of L-DOPA

Each T-75 flask was transfected with 9 µg of GFP_40TAG, 4.5 µg of H1-1bp_tRNACUA, and 4.5 µg of L-DOPA_39CP1. 12-24 hours after transfection the growth media of the cells was replaced with one containing: 4% FBS, 1X MEM Non-Essential Amino Acids (Sigma Aldrich, St. Louis, MO), 1X Antibiotic/Antimycotic, 5 mM L-DOPA, and 10 mM sodium ascorbate. Approximately 96 hours after transfection, the HEK293T cells were washed 2X with phosphate buffered saline (Sigma Aldrich, St. Louis, MO) and collected by centrifugation. Some cells were fixed in a phosphate buffered saline containing 4% paraformaldehyde in (USB Corp., Cleveland, OH) and stored at 4 ºC for later FACS analysis. Most were rapidly frozen and stored at -80 ºC until purification and detection.

FACS Analysis

A BD FACSCalibur flow cytometer was used to gate approximately 10,000 cells based on forward and side scatter. Each cell was then excited with 488 nm light and its resulting emission detected with an FL-1 filter (515-545 nm). Positive fluorescent populations were gated based on negative controls or positive controls and analyzed using the computer software Cyflogic v.1.2.1 (CyFlo, Ltd., Turku, Finland).

Purifying and Detecting L-DOPA Incorporated GFP

Approximately 6 g of frozen HEK293T cells (-80 ºC) were thawed and resuspended in 30 mL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 0.05% Tween-20, 6 mM sodium ascorbate, pH 8.0). A complete protease cocktail inhibitor (Roche, Indianapolis, IN) was also added. Cells were then lysed by sonification and centrifuged at 14,000 rpm for 25 minutes at 4 ºC. The supernatant was

applied to Ni-NTA agarose resin (Qiagen, Madison, WI) for two hours at 4 ºC allowing 6XHis tagged GFP to be bound. The resin was then washed with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 30 mM imidazole, 10 mM sodium ascorbate, pH 8.0). L-DOPA incorporated GFP was then eluted from the beads with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0). A Nanocep 10K Omega centrifugal device (Pall Corp., Port Washington, NY) was used to concentrate the eluted protein. Concentrated protein was resolved by SDS-PAGE using a 14% Tris-glycine polyacrylamide gel (Invitrogen, Calsbad, CA) in SDS running buffer (25 mM Tris base, 192 mM glycine, 0.1% SDS, pH 8.3). Proteins were then transferred onto an Immobilon PVDF membrane (Millipore) in transfer buffer (48 mM Tris base, 39 mM glycine, 20% methanol (v/v), pH 9.2) using a BioRad Semi-Dry Blotter. The membrane was then probed for full length GFP expression using a monoclonal anti-EGFP antibody (Clontech, Mountain View, CA) and a goat anti-mouse polyclonal antibody conjugated with horseradish perodixase (Bio-Rad, Hercules, CA). To visualize the bands, ECL Western Blotting Detection Reagents (GE Healthcare, Sweden) was applied to the membrane for 5 minutes and exposed to a BioMax light film (Eastman Kodak Co., Rochester, NJ). Staining of L-DOPA incorporated GFP was carried by first rinsing the membrane with deionized water after transferring and stained in staining solution (2 M sodium glycinate, 0.24 mM nitroblue tetrazolium, pH 10.0) on a rotary shaker in darkness overnight.

4.2 WORKING TOWARDS THE INCORPORATION OF OTHER NONNATURAL AMINO ACIDS IN MAMMALIAN CELLS

4.2.1 Introduction

The advantage of using the connective polypeptide 1 (CP1) transplantation approach to develop orthogonal pairs for incorporation is that it allows us to rapidly shuttle previously evolved nonnatural aaRS into mammalian systems. Several nonnatural aaRS have been evolved from *M. jannaschii* TyrRS. These mutant enzymes, when paired with an orthogonal suppressor tRNA have been able to incorporate a wide variety of nonnatural amino acids in proteins in *E. coli* cells with specificities > 99% and efficiencies up to 75% of wild type (tyrosine) incorporation [5]. Because both the *M. jannaschii* derived tRNA and mutant TyrRS cross-react with *eukaryotic* tyrosyl-tRNA and TyrRS, these pairs have not been used in mammalian systems.

It is very difficult create a nonnatural aaRS using rationale design. Instead, mutants are screened from large genetic libraries (10^9 variations) using alternating rounds of positive and negative selection [6]. The only mammalian cells useable nonnatural aaRS have been created by evolving an *E. coli* TyrRS in yeast cells [7]. These evolved aaRS can be used in mammalian cells. However, yeast genetic libraries are smaller in size and more difficult to screen than *prokaryotic* ones. Our approach would allow the existing mutant *M. jannaschii* TyrRS, evolved in *E. coli*, to be used in mammalian cells. The previous section focused on our efforts to incorporate 3,4-dihydroxy-L-phenylalanine (L-DOPA) using a previously evolved L-DOPA_RS derived from *M. jannaschii* TyrRS. Our final efforts will concentrate on manipulating a couple of other *M. jannaschii* TyrRS evolved nonnatural aaRS. This section will focus on our efforts to shuttle the previously evolved pair for *p*-iodo-L-phenylalanine and *p*-acetyl-L-phenylalanine into the mammalian system (**Figure 4.4**).

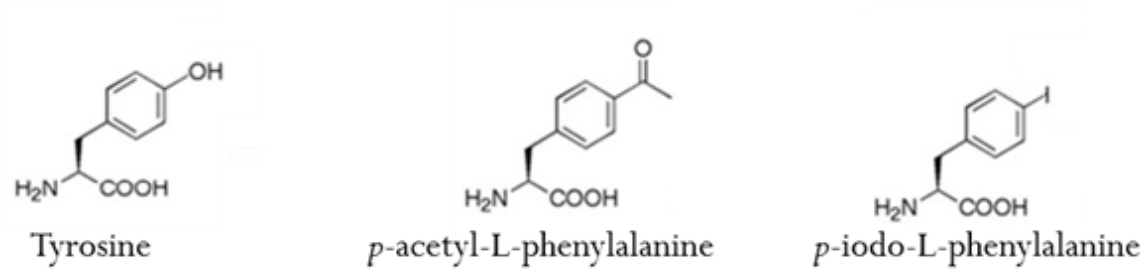


Figure 4.4: The natural amino acid tyrosine and two nonnatural analogs. In order for a mutant aaRS to recognize the nonnatural amino acids, its active site must be mutated to accommodate the difference in side chain structure. Both ketone and iodo functional groups are not found on any of the twenty natural amino acids. Both have been successful incorporated into mammalian cells using an *E. coli* derived pair [8] and into *prokaryotic* cells using a *M. jannaschii* derived pair [9, 10].

4.2.2 Results and Discussion

Xie *et al.* previously evolved a nonnatural aaRS with specificity for *p*-iodo-L-phenylalanine (p-iodo). This mutant *M. jannaschii* TyrRS contained five mutations in its active site that are responsible for recognizing the amino acid (Leu32, Ser107, Pro158, Leu159 and Glu162) [10]. As discussed before, their *M. jannaschii* derived pair is not orthogonal in mammalian cells. Since a CP1 substitution was able to change the orthogonality (tRNA specificity) of *M. jannaschii* TyrRS we season that a similar approach could be done to the P-IODO specific aaRS. The *E. coli* TyrRS's CP1 domain was substituted into the P-IODO aaRS to make P-IODO_39CP1. We hope this would change its tRNA specificity from recognizing a C1:G72 sequence to a G1:C72. As a result, it would no longer recognize endogenous tyrosyl-tRNA's but can charge a 1bp-$tRNA_{CUA}$, completing an orthogonal pair.

We tested the incorporation efficiency of this pair by using a functional assay (previously described). A GFP gene harboring an amber stop codon, GFP_40TAG, was transfected into HEK293T cells along with H1_1bp-tRNA$_{CUA}$ and P-IODO_39CP1. 24 hours later P-IODO was added to the growth medium. 96 hours after transfection the cells were harvested and full length GFP expression was detected using FACS (**Figure 4.5**).

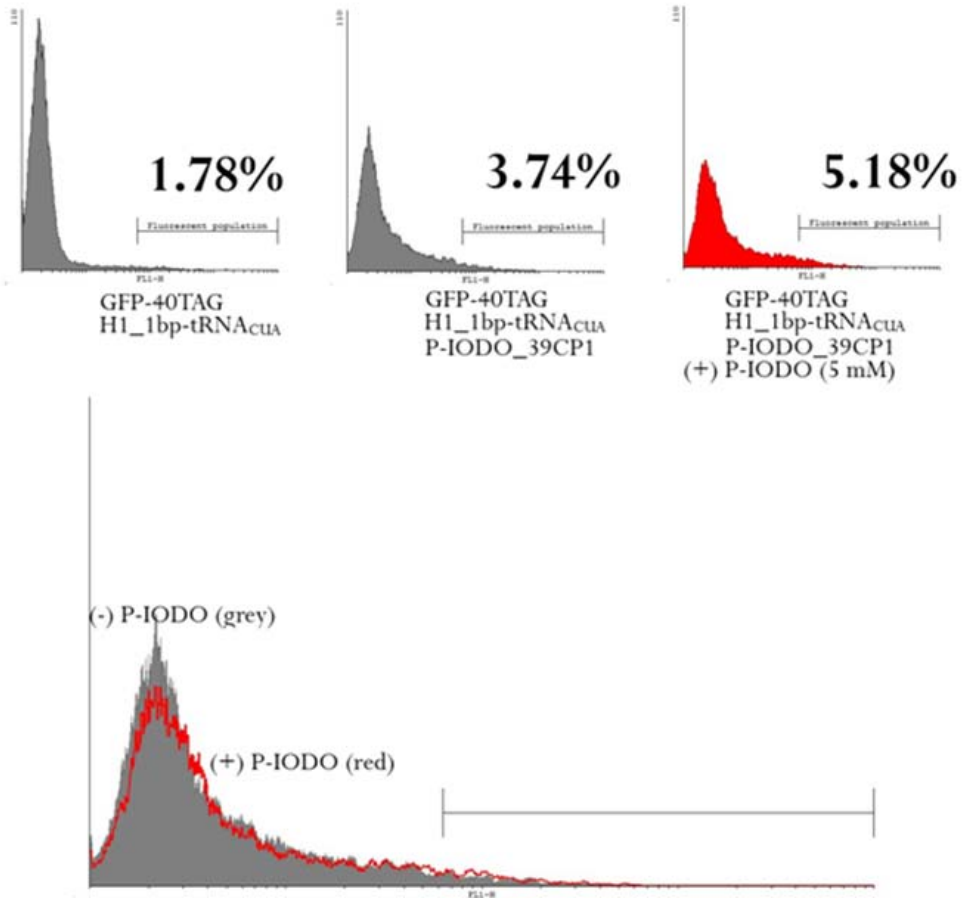


Figure 4.5: Measuring the amount of fluorescent cells using FACS. The P-IODO_39CP1 was able to charge 1bp-tRNACUA in the absence of P-IODO. Most likely, it was still able to recognize tyrosine. Addition of P-IODO caused a <1.5% increase in fluorescing cells. This suggests that P-IODO_39CP1 was able to recognize P-IODO and charge it to 1bp-tRNA$_{CUA}$.

As in the case of L-DOPA incorporation, when all three plasmids were present but P-IODO not supplied we still saw some increase in full length GFP expression. This is most likely due to charging of 1bp-$tRNA_{CUA}$ with a natural amino acid such as tyrosine. However, when P-IODO was added we saw a 2% increases in suppression efficiency. This could be due to a P-IODO charged 1bp-$tRNA_{CUA}$.

We also applied our CP1 methodology to a previously evolved nonnatural aaRS with specificity for *p*-acetyl-L-phenylalanine (P-KETO). The P-KETO specific aaRS was also evolved from *M. jannaschii* TyrRS and contains the following mutations in its active site: Leu32, Gly158, Cys159, and Arg162 [9]. Other groups have shown that this P-KETO aaRS has very high fidelity (>99.8%) and activity. It is able to yield of more than >3.6 mg of incorporated protein per 1 L of *E. coli* culture. We applied our CP1 transplantation to the P-KETO aaRS to make P-KETO_39CP1. The pair was tested using a similar assay that was used to test P-IODO (**Figure 4.6**). It seems that the P-KETO_39CP1, just like P-IODO_39CP1 and L-DOPA_39CP1, is able to still recognize natural amino acids in the absence of the nonnatural amino acid. Addition of P-KETO caused a very small increase in the population of fluorescing cells. As in the case with the P-IODO results, FACS cannot tell us the identity of the incorporated amino acid so additional assay will have to be done.

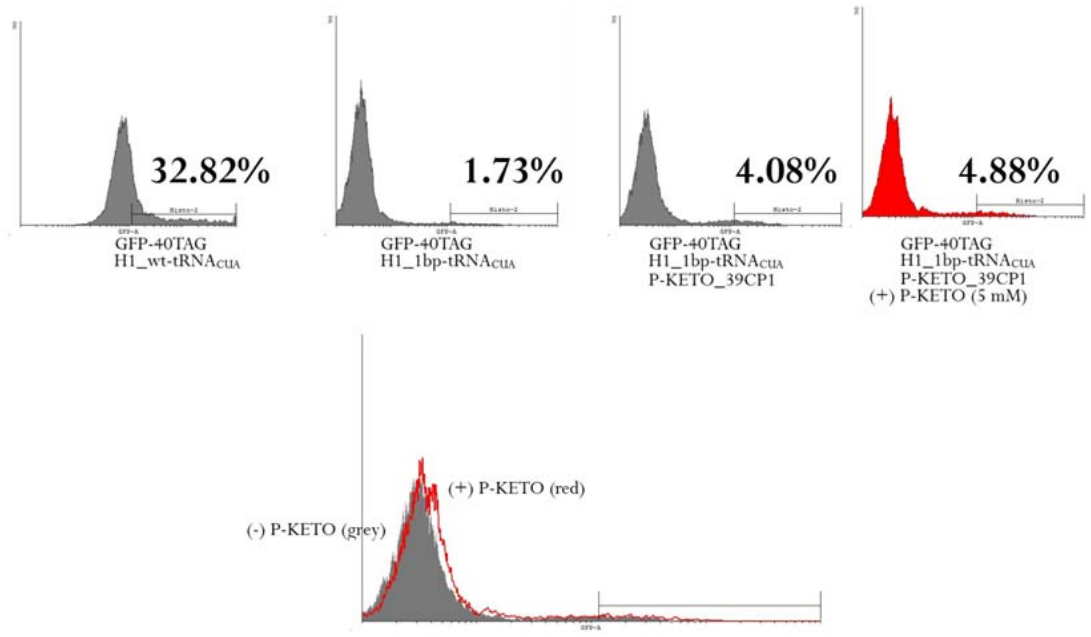


Figure 4.6: Measuring the amount of fluorescent cells using FACS. The P-KETO_39CP1 was able to charge 1bp-tRNA$_{CUA}$ in the absence of the nonnatural amino acid. Adding P-KETO resulted in a <1% increase in fluorescing cells, possible due to incorporation of P-KETO.

This section demonstrated our application of the CP1 substitution on two different nonnatural aaRS. In both cases, we observed a suspected drop in amino acid fidelity. The CP1 substitution might be allowing it to recognize some natural amino acids such as tyrosine. However, when the nonnatural amino acid was present, we did see more charged 1bp-tRNA$_{CUA}$. Subsequent experiments will have to be conducted to verify the identity of the amino acid on 1bp-tRNA$_{CUA}$. Other nonnatural aaRS can also be CP1 swapped and tested to see if they maintain their functionality and specificity as a result of the substitution.

4.2.3 Materials and Methods

Cell Culturing and Transfection of HEK293T Cells

Plasmids were amplified by transforming electrically component *E. coli* TOP10 cells, grown in LB, and isolated using QIAprep Spin Miniprep Kit (Qiagen, Madison, WI). HEK293T cells (ATCC CRL-11268) were cultured in Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% (v/v) fetal bovine serum (Atlanta Biologicals, #C0136) at 37 ºC in a humidified incubator containing 5% CO_2. Transfections complexes were formed using FuGENE HD (Roche, Indianapolis, IN). Plasmids were diluted in OPTI-MEM (Invitrogen, Carlsbad, CA) and FuGENE HD was added at a 2:1 (µL FuGENE HD: µg DNA). After 15 minutes, this was added to HEK293T cells (confluency 40-60%) in T-75 culture flasks.

Plasmid Construction

The construction of the GFP_40TAG and H1-1bp-$tRNA_{CUA}$ plasmids is described in previous chapters. P-KETO_39CP1 was generated by replacing amino acid 110-148 in the P-KETO specific aaRS with amino acids 129-167 from *E. coli* TyrRS. The 5' segment (segment 1) of a previously evolved P-KETO aaRS gene was amplified from the plasmid containing the mutant gene (Gift from Dr. Peter G. Schultz, The Scripps Research Institute, La Jolla, CA) with the following primers:

5'-GCGGATCCGCCACCATGGACGAATTTGAAATGATAAAGAGAA'3

5'-ACATTCATATTGCCGAACCAATAAACATATTTTGCCTT-'3

A sequence (segment 2) for *E. coli* TyrRS's CP1 domain was amplified by PCR using:

5'-AAGGCAAAATATGTTTATTGGTTCGGCAATATGAATGT-'3

5'-ATAACTTCAGCAACCTTTGGCCCCTGATCTTCACGGTTGA-'3

Finally, the 3’ fragment (segment 3) of the P-KETO aaRS was amplified from plasmid using:

5’-TCAACCGTGAAGATCAGGGGCCAAAGGTTGCTGAAGTTAT-‘3

5’-GCTCTAGAGCTTATAATCTCTTTCTAATTGGCTCTAAAATCTTTATA

AGTTCTTCAGCTACAGCATTTTTTAACCTCATTGGATGCAATTCCTT-‘3

The segment 1 forward and segment 2 reverse primers were used to amplify a sequence containing segments 1 and 2. Finally, this segment 1_2 joined to segment 3 by PCR using segment 1 forward and segment 3 reverse primers. This gene also contained a Kozak initiation sequence and the Arg286 mutation. The segment 1_2_3 PCR product was digested with BamHI and XbaI (New England Biolabs, Ipswich, MA) and ligated using DNA T4 Ligase into a predigested pEF6-V5 (Invitrogen, Carlsbad, CA) to create P-KETO_39CP1. P-IODO_39CP1 was generated in a similar manner, using the same primers. A P-IODO aaRS, evolved from *M. jannaschii* TyrRS, was used as a template (Gift from Dr. Peter G. Schultz, The Scripps Research Institute, La Jolla, CA).

Genetic Incorporation of P-IODO and P-KETO

HEK293T cells were transfected in six-well plates (2 mL media each). 24 hours later, the media was supplement with the nonnatural amino acid at 5 mM. After 72 hours, the cells were harvested, fixed in PBS containing 4% paraformaldehyde in (USB Corp., Cleveland, OH) and stored at 4 ºC for FACS analysis. Both P-KETO (Chem-Impex, Dale, IL) and P-IODO (Fischer Scientific, Pittsburgh, PA) were purchased commercially.

FACS Analysis

A BD FACSCalibur flow cytometer was used to gate approximately 10,000 cells based on forward and side scatter. Each cell was then excited with 488 nm light and its

resulting emission detected with an FL-1 filter (515-545 nm). Positive fluorescent populations were gated based on negative controls or positive controls and analyzed using the computer software Cyflogic v.1.2.1 (CyFlo, Ltd., Turku, Finland).

4.3 CONCLUSION

The development of orthogonal nonnatural suppressor tRNA/aaRS pairs is a challenging and tedious process. Our objective was to outline an approach that can be used to rapidly create nonnatural tRNA/aaRS pairs for use in mammalian systems. Dozens of these pairs have already been evolved from *M. jannaschii* tyrosyl-tRNA/TyrRS and can be used use in *prokaryotic* systems. Our approach involves the manipulation of the orthogonality of these pairs to make them applicable. We did this in a three step approach. Chapter 2 discussed our efforts to switch the orthogonality of a *M. jannaschii* tyrosyl-tRNA derived suppressor tRNA. Our efforts resulted in the proficient expression of a functional and orthogonal amber codon suppressing tRNA (1bp-$tRNA_{CUA}$). Chapter 3 centered on the manipulation of the corresponding *M. jannaschii* TyrRS. A CP1 substitution was used to change the TyrRS's recognition of a C1:G72 containing tRNA to a G1:C72. As a result, it should be orthogonal in mammalian cells since it could no longer recognize endogenous tyrosyl-tRNAs that contain a C1:G72. Furthermore, the CP1 substitution gave it the ability it to recognize and charge our G1:C72 containing tRNA.

In this final chapter, we completed our nonnatural tRNA/aaRS pair by applying the CP1 substitution to some mutant *M. jannaschii* TyrRS. Three previously evolved nonnatural aaRS were manipulated with the CP1 substitution and tested for their ability to charge 1bp-$tRNA_{CUA}$ with their respective nonnatural amino acid. We then combined

the nonnatural CP1 swapped aaRS with our orthogonal suppressor tRNA to form an orthogonal pair. We investigated the ability of these three pairs to site-specifically incorporate a nonnatural amino acid in response to an amber stop codon in a gene. Efforts to incorporate 3,4-dihydroxy-L-phenylalanine (L-DOPA) yielded promising results but will require optimization. The incorporation of P-IODO and P-KETO might have been successful but is yet unconfirmed. Our low incorporation efficiency in mammalian cells is lower than the observed incorporation in *prokaryotic* cells (using a NON-CP1 swapped *M. jannaschii* pair). It is almost certain that mutating the aaRS's specificity for both its amino acid and tRNA will negatively impact its activity. Depending on the already existing active site mutations, the additional CP1 substitution may be too much of a change to maintain the specificity and functionality of the synthetase. However, we believe that our approach can be tested with other nonnatural aaRS that were also evolved from *M. jannaschii* TyrRS and evaluated on a case by case basis.

4.4 REFERENCES

1. Wang, L. and P.G. Schultz, *Expanding the genetic code.* Angewandte Chemie-International Edition, 2005. **44**(1): p. 34-66.
2. Umeda, A., et al., *Site-specific Protein Cross-Linking with Genetically Incorporated 3,4-Dihydroxy-L-Phenylalanine.* Chembiochem, 2009. **10**(8): p. 1302-1304.
3. Alfonta, L., et al., *Site-specific incorporation of a redox-active amino acid into proteins.* J Am Chem Soc, 2003. **125**(48): p. 14662-3.
4. Fluckiger, R., T. Woodtli, and P.M. Gallop, *The interaction of aminogroups with pyrroloquinoline quinone as detected by the reduction of nitroblue tetrazolium.* Biochem Biophys Res Commun, 1988. **153**(1): p. 353-8.
5. Wang, L., et al., *Expanding the genetic code of Escherichia coli.* Science, 2001. **292**(5516): p. 498-500.
6. Liu, D.R., et al., *Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo.* Proc Natl Acad Sci U S A, 1997. **94**(19): p. 10092-7.
7. Chin, J.W., et al., *An expanded eukaryotic genetic code.* Science, 2003. **301**(5635): p. 964-7.
8. Ye, S., et al., *Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis.* J Biol Chem, 2008. **283**(3): p. 1525-33.

9. Wang, L., et al., *Addition of the keto functional group to the genetic code of Escherichia coli.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 56-61.

10. Xie, J., et al., *The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination.* Nat Biotechnol, 2004. **22**(10): p. 1297-301.

Appendix

Figure A.1: CP1 substitutions for the various TyrRS mutations. Column 2 shows the amino acids removed from *M. jannaschii* TyrRS and replaced with amino acids from either *E. coli* TyrRS or *T. thermophilus* TyrRS.

Plasmid	Removed (AA)	Added from E. coli TyrRS (AA)	Added from T. thermophilus TyrRS (AA)
TyrRS_36CP1	110-145	129-167	
TyrRS_39CP1	105-143	129-167	
TyrRS_39CP1_2	110-148	129-167	
TyrRS_42CP1	101-134	122-163	
TyrRS_44CP1	110-148	129-172	
TyrRS_fullCP1	105-162	129-191	
TyrRS_29RED	110-138	129-162	
TyrRS_34RED	105-138	129-162	
TyrRS_38RED	101-138	122-162	
TyrRS_39tt	105-143		129-167
TyrRS_45tt	99-143		123-167

Figure A.2: FACS analysis using CyFlogic of acceptor stem modifications. Well A: GFP_40TAG, Well B: GFP_40TAG and H1-wt-tRNA$_{CUA}$, Well C: GFP_40TAG and H1-1bp-tRNA$_{CUA}$, Well D: GFP_40TAG and H1-2bp-tRNA$_{CUA}$, Well E: GFP_40TAG and H1-3bp-tRNA$_{CUA}$, Well F: GFP

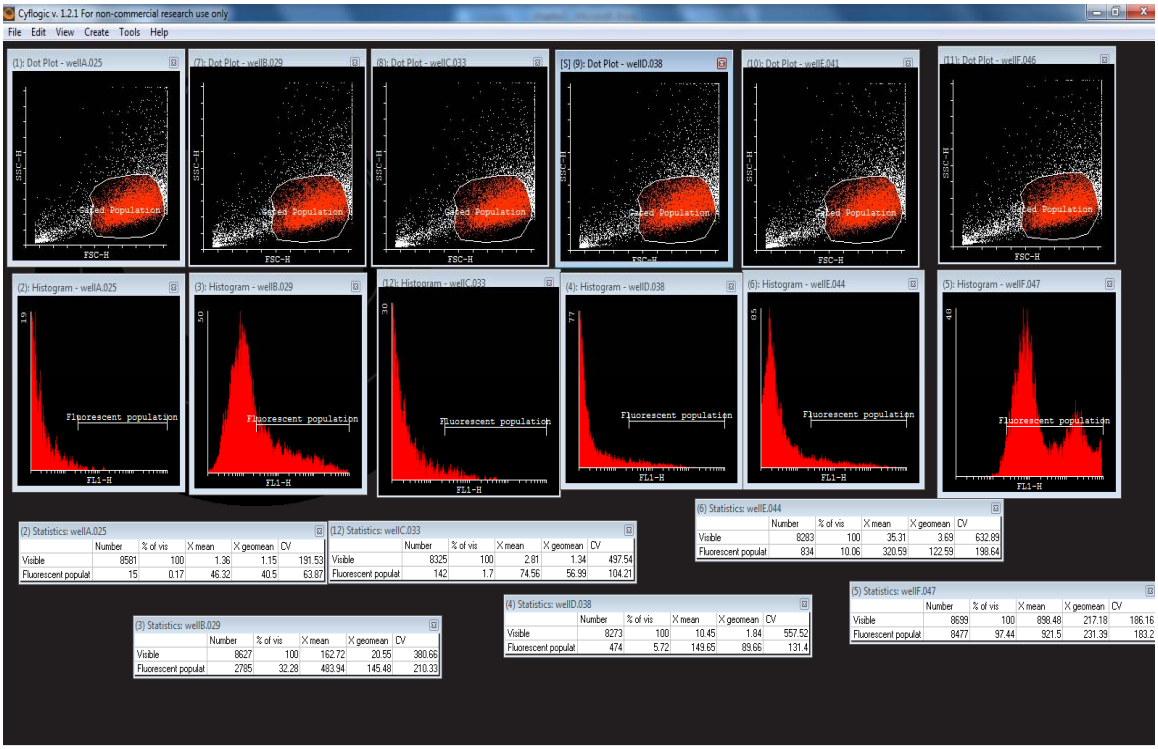


Figure A.3: FACS analysis using Cyflogic of *E. Coli* TyrRS charging 1bp-tRNA$_{CUA}$. Tube 7: GFP_40TAG, Tube 8: GFP_40TAG, H1_wt-tRNA$_{CUA}$, Tube 9: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, Tube 10: GFP_40TAG, H1_wt-tRNA$_{CUA}$, *E. coli* TyrRS

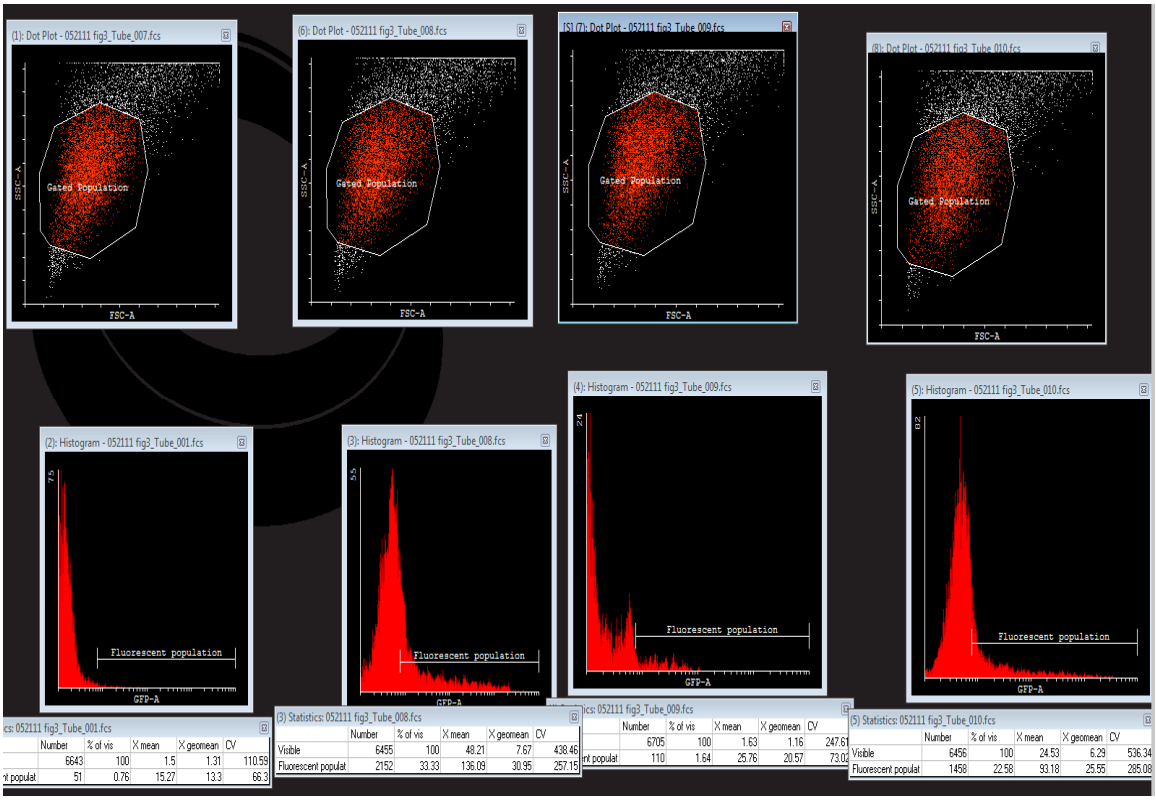


Figure A.4: FACS analysis using Cyflogic of TyrRS_44CP1 charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, 6x_1bp-tRNA$_{CUA}$, Right: GFP_40TAG, 6x_1bp-tRNA$_{CUA}$, TyrRS_44CP1

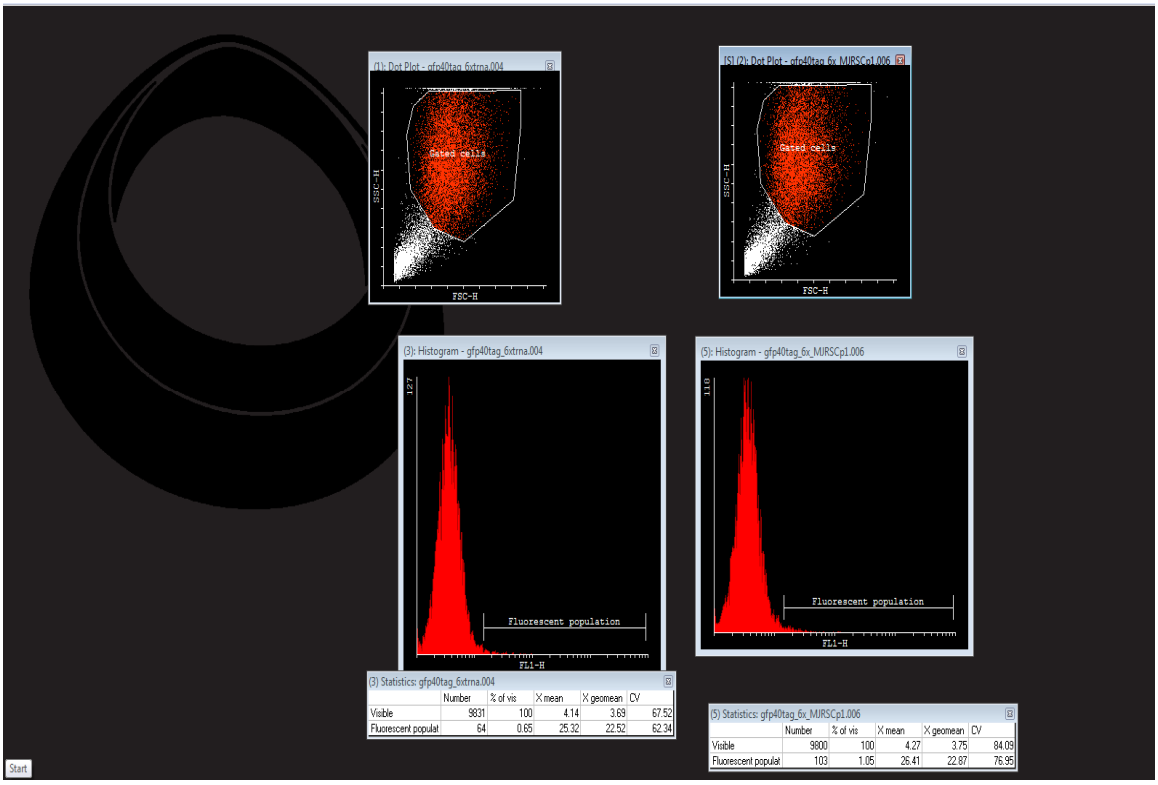


Figure A.5: FACS analysis using Cyflogic of TyrRS_44CP1_R charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, Middle: GFP_40TAG, 6x_1bp-tRNA$_{CUA}$, Right: GFP_40TAG, 6x_1bp-tRNA$_{CUA}$, TyrRS_44CP1_R

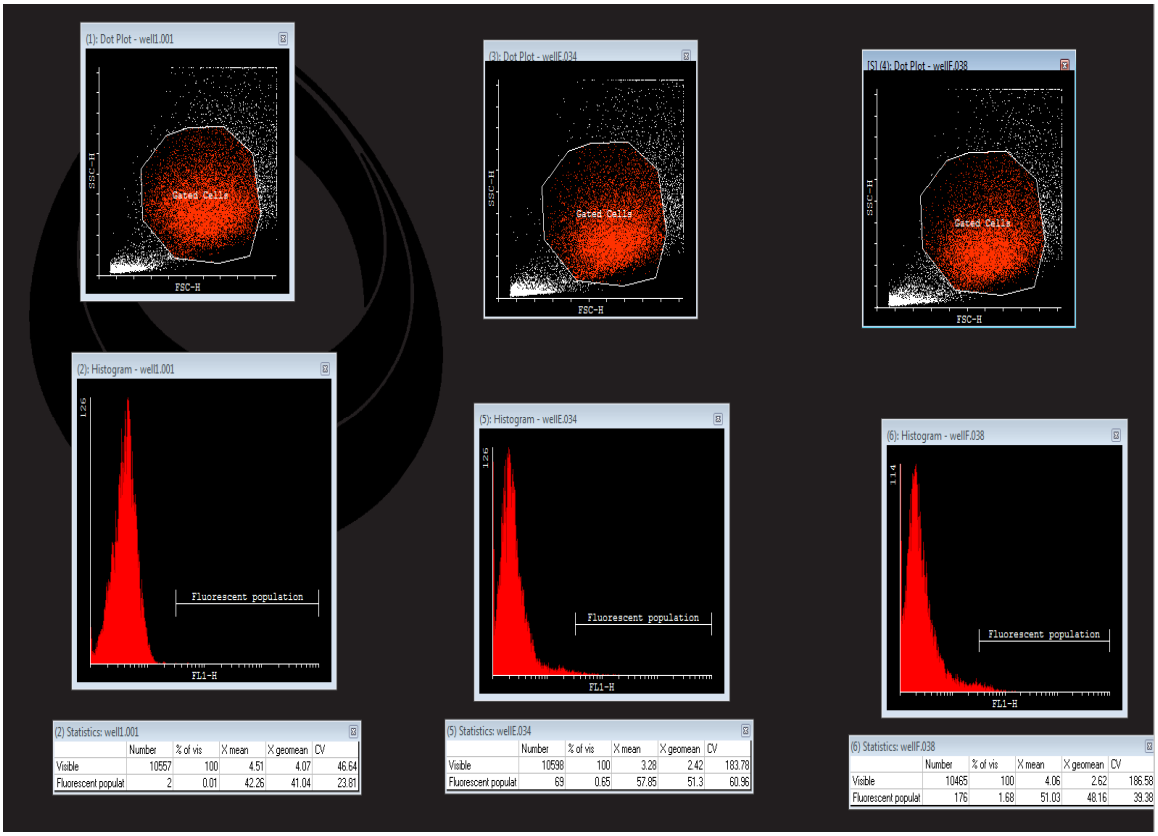


Figure A.6: FACS analysis using Cyflogic of TyrRS_44CP1_R charging 1bp-tRNA$_{CUA}$. Left: GFP, Left-Middle: GFP_40TAG, Right-Middle: GFP_40TAG H1_1bp-tRNA$_{CUA}$, Right: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_44CP1_R

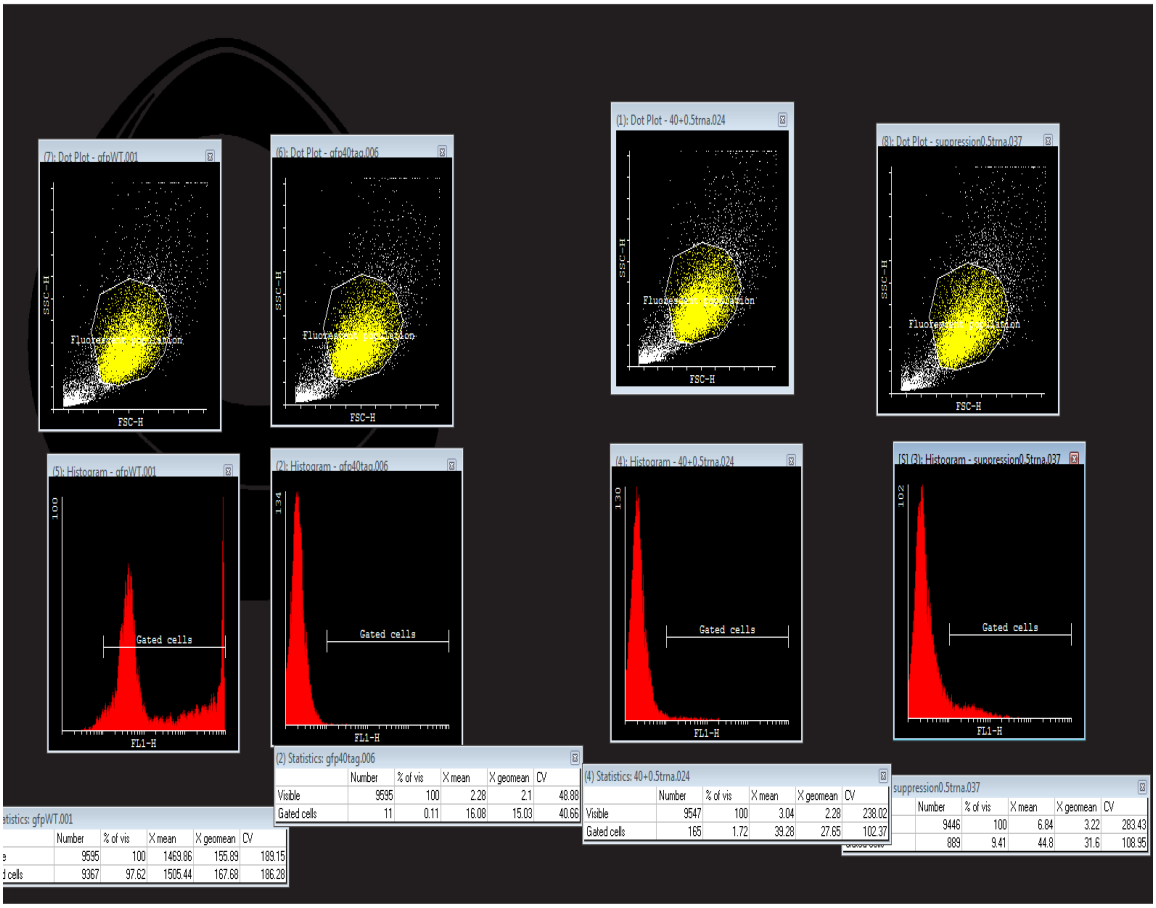


Figure A.7: FACS analysis using Cyflogic of TyrRS_39CP1_2 charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, Left-Middle: GFP_40TAG, H1_wt-tRNA$_{CUA}$ Right-Middle: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, Right: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_39CP1_2

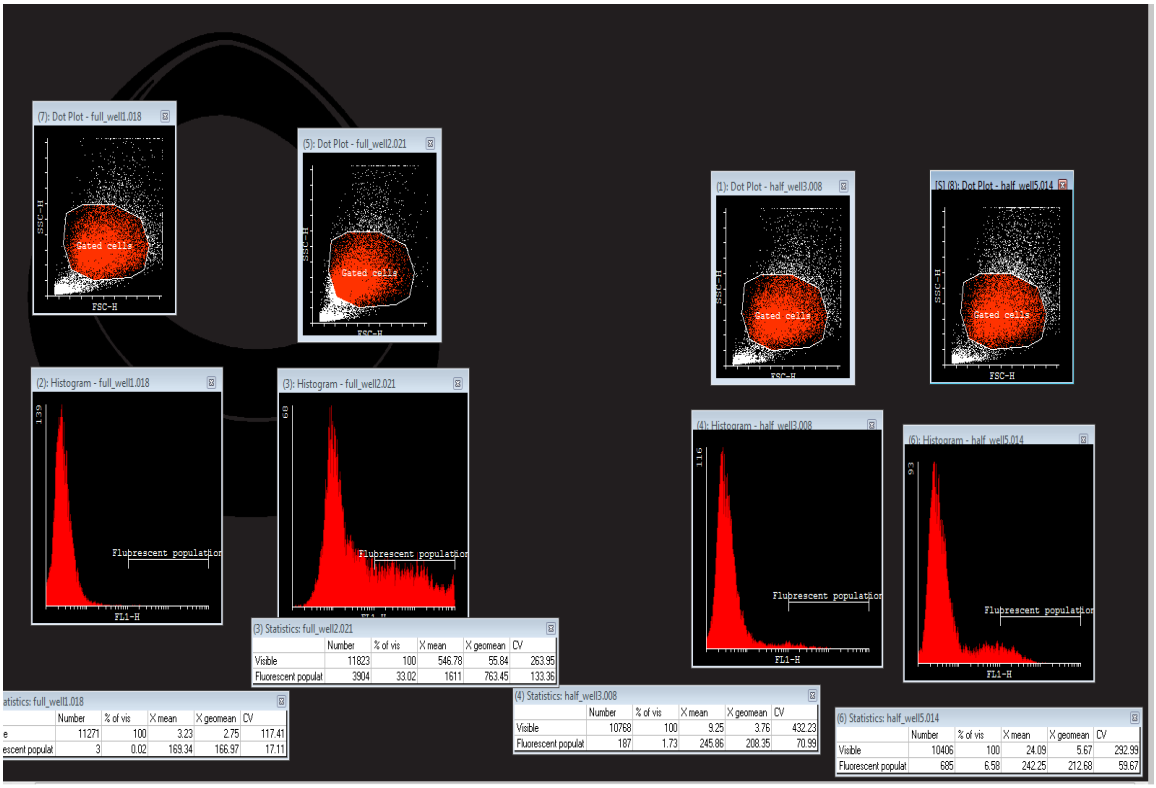


Figure A.8: FACS analysis using Cyflogic of TyrRS_36CP1 charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, Right: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_36CP1_2

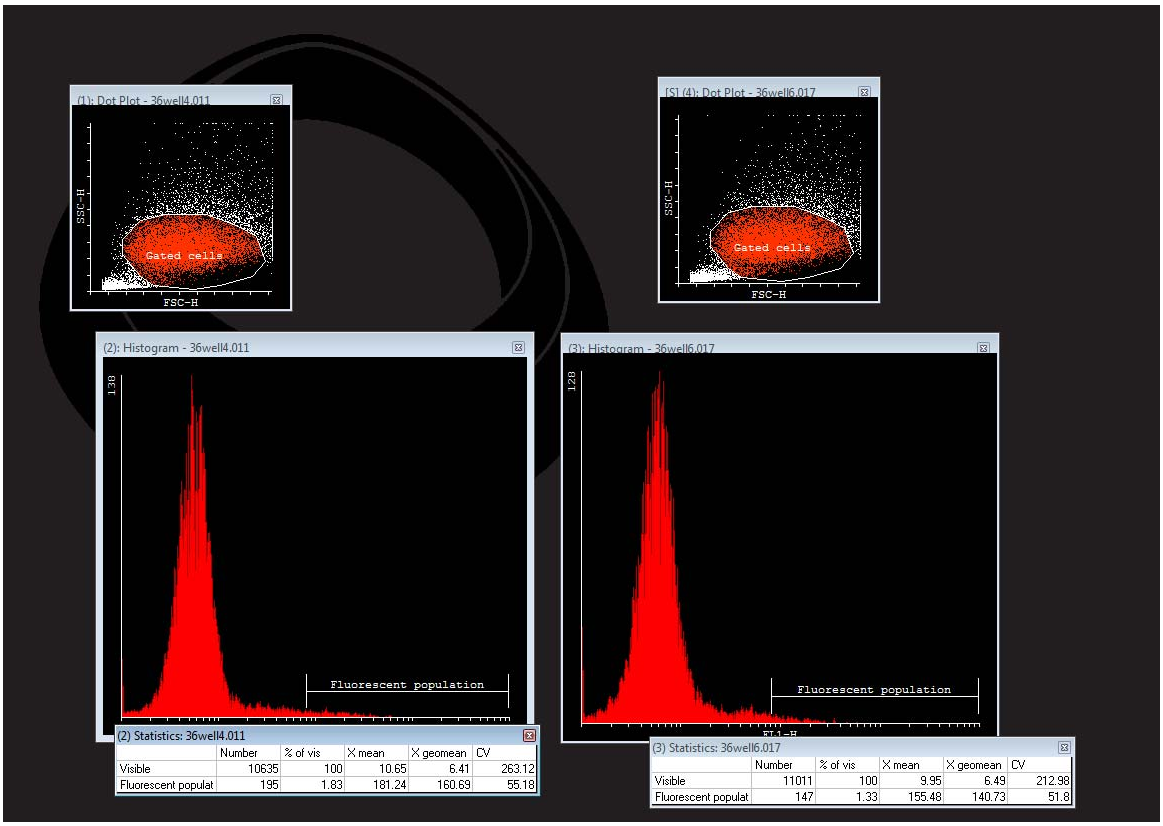


Figure A.9: FACS analysis using Cyflogic of RED CP1 substituted mutants charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, Left-Middle: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_29RED, Right-Middle: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_34RED, Right: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_38RED

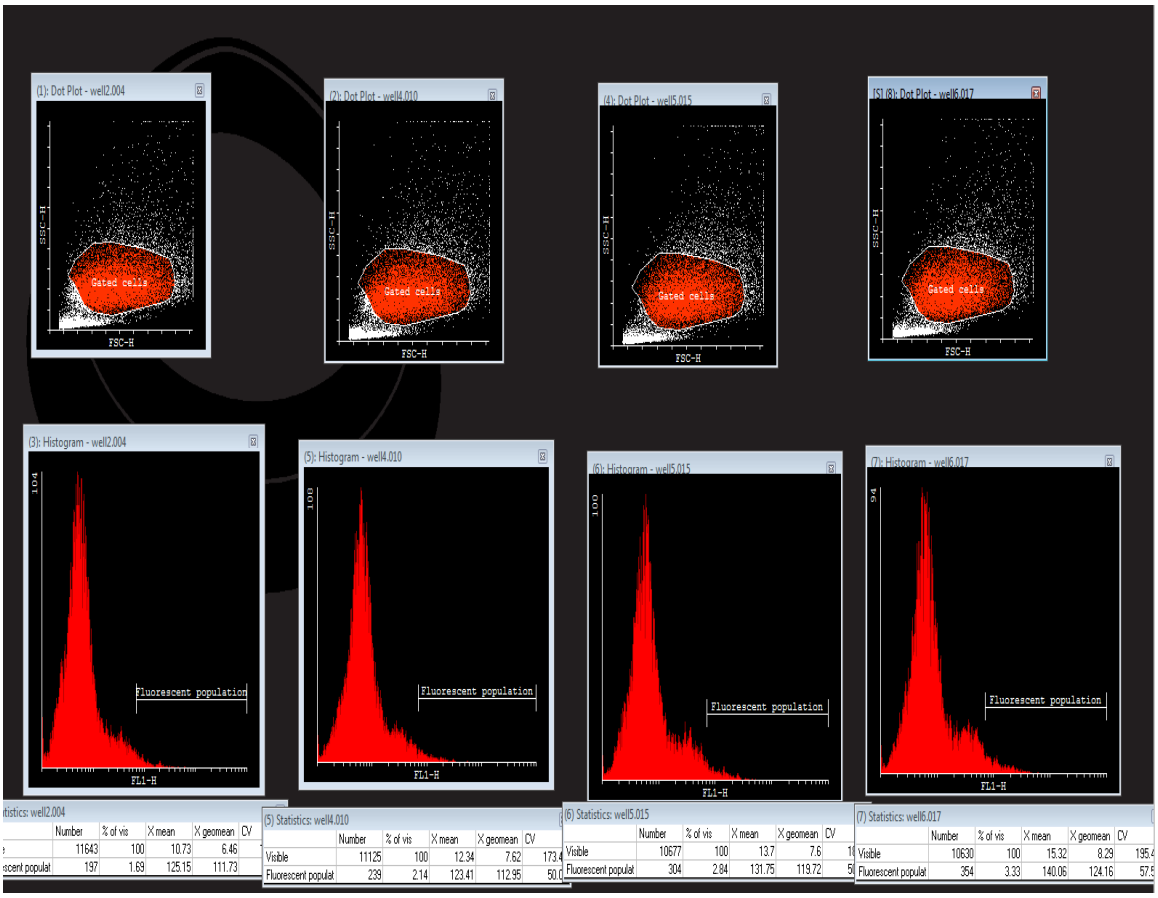


Figure A.10: FACS analysis using Cyflogic of *T. thermophilus* CP1 substituted mutants charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, Middle: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_39tt, Right: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_45tt

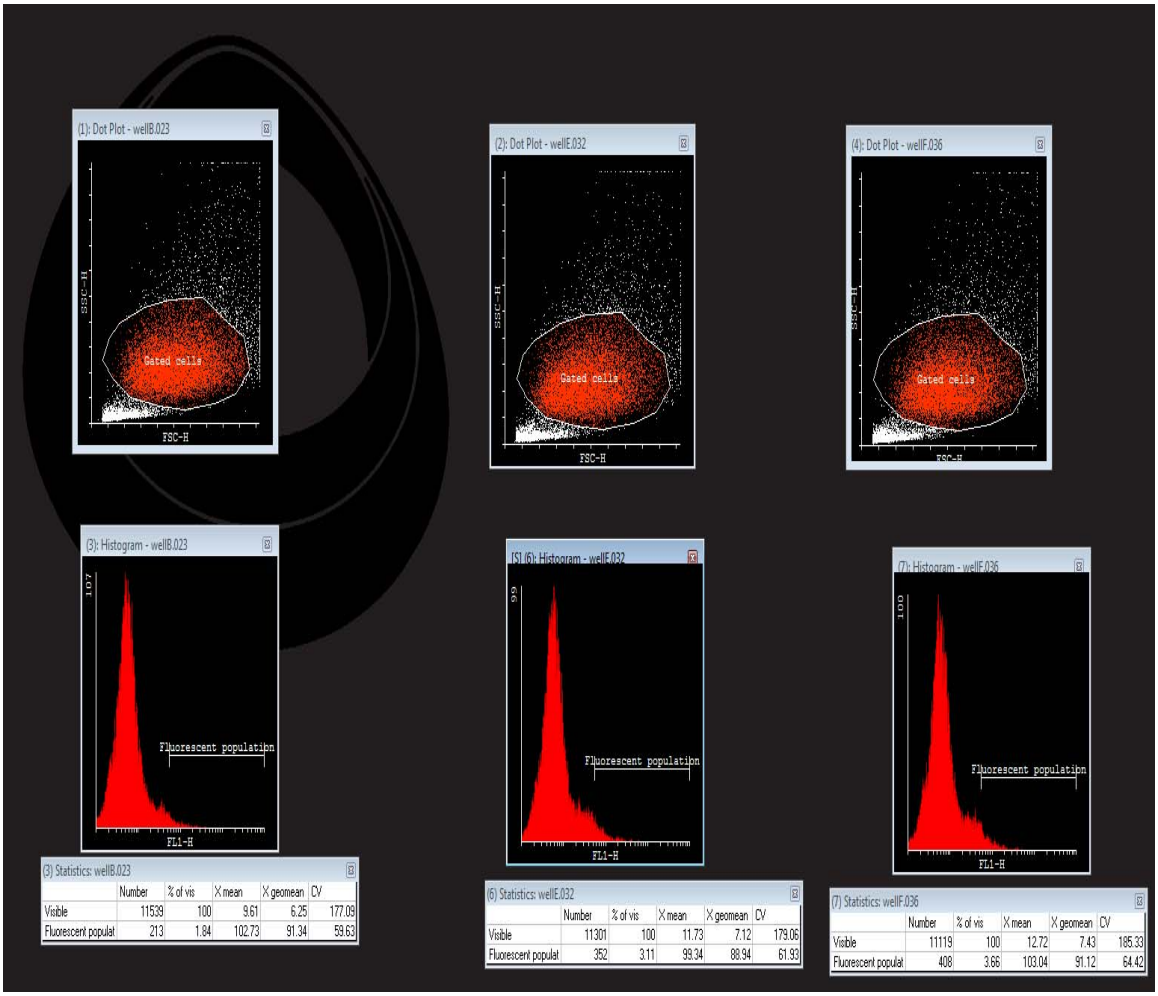


Figure A.11: FACS analysis using Cyflogic of TyrRS_fullCP1 charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, Right: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, TyrRS_fullCP1

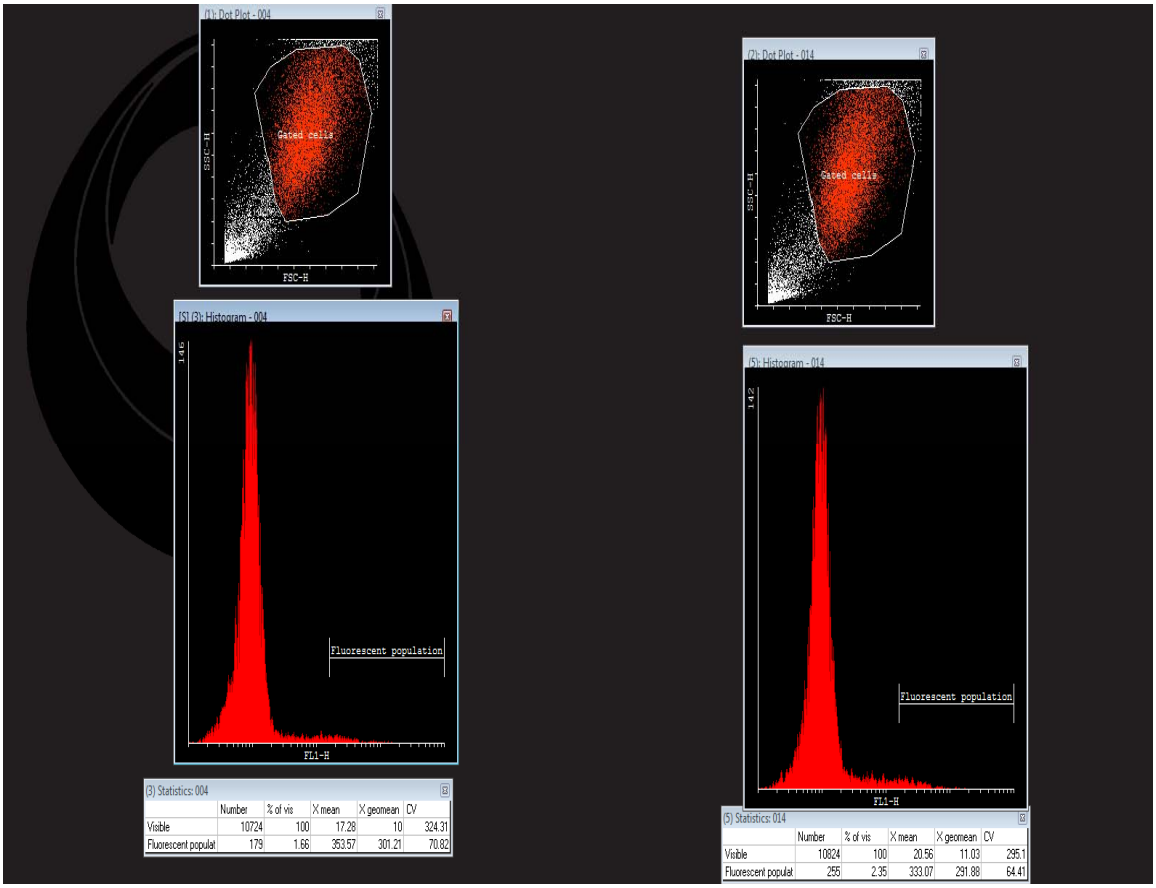
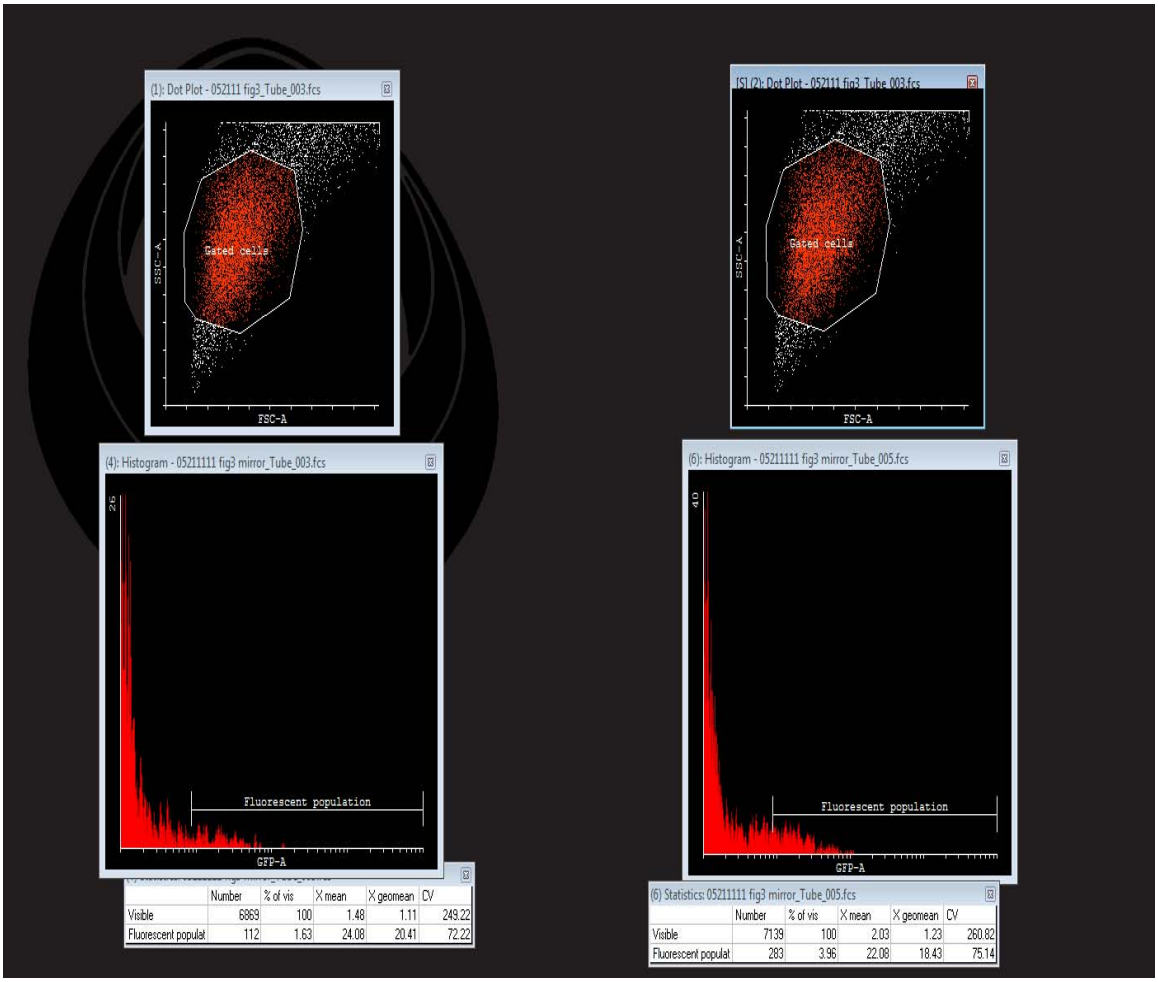


Figure A.12: FACS analysis using Cyflogic of *M. jannaschii* TyrRS charging 1bp-tRNA$_{CUA}$. Left: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, Right: GFP_40TAG, H1_1bp-tRNA$_{CUA}$, *M. jannaschii* TyrRS



Bibliography

Alfonta, L., Z. Zhang, et al. (2003). "Site-specific incorporation of a redox-active amino acid into proteins." J Am Chem Soc **125**(48): 14662-3.

Bain, J. D., E. S. Diala, et al. (1989). "Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide." Journal of the American Chemical Society **111**(20): 8013-8014.

Beatty, K. E. and D. A. Tirrell (2008). "Two-color labeling of temporally defined protein populations in mammalian cells." Bioorg Med Chem Lett **18**(22): 5995-9.

Bose, M., D. Groff, et al. (2006). "The incorporation of a photoisomerizable amino acid into proteins in E. coli." J Am Chem Soc **128**(2): 388-9.

Budisa, N., C. Minks, et al. (1998). "Residue-specific bioincorporation of non-natural, biologically active amino acids into proteins as possible drug carriers: structure and stability of the per-thiaproline mutant of annexin V." Proc Natl Acad Sci U S A **95**(2): 455-9.

Calendar, R. and P. Berg (1966). "The catalytic properties of tyrosyl ribonucleic acid synthetases from Escherichia coli and Bacillus subtilis." Biochemistry **5**(5): 1690-5.

Carter, P. (1986). "Site-directed mutagenesis." Biochem J **237**(1): 1-7.

Chin, J. W., T. A. Cropp, et al. (2003). "An expanded eukaryotic genetic code." Science **301**(5635): 964-7.

Chin, J. W., T. A. Cropp, et al. (2003). "An expanded eukaryotic genetic code." Science **301**(5635): 964-7.

Cohen, G. N. and R. Munier (1956). "[Incorporation of structural analogues of amino acids in bacterial proteins]." Biochim Biophys Acta **21**(3): 592-3.

Cowie, D. B., G. N. Cohen, et al. (1959). "Amino acid analog incorporation into bacterial proteins." Biochim Biophys Acta **34**: 39-46.

Dawson, P. E., T. W. Muir, et al. (1994). "Synthesis of proteins by native chemical ligation." Science **266**(5186): 776-9.

Deiters, A., T. A. Cropp, et al. (2004). "Site-specific PEGylation of proteins containing unnatural amino acids." Bioorg Med Chem Lett **14**(23): 5743-5.

Deutscher, M. P. (1984). "Processing of tRNA in prokaryotes and eukaryotes." CRC Crit Rev Biochem **17**(1): 45-71.

Dieterich, D. C., A. J. Link, et al. (2006). "Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT)." Proc Natl Acad Sci U S A **103**(25): 9482-7.

Dougherty, D. A. (2000). "Unnatural amino acids as probes of protein structure and function." Current Opinion in Chemical Biology **4**(6): 645-652.

Fluckiger, R., T. Woodtli, et al. (1988). "The interaction of aminogroups with pyrroloquinoline quinone as detected by the reduction of nitroblue tetrazolium." Biochem Biophys Res Commun **153**(1): 353-8.

Garnier, J., D. J. Osguthorpe, et al. (1978). "Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins." J Mol Biol **120**(1): 97-120.

Giege, R., M. Sissler, et al. (1998). "Universal rules and idiosyncratic features in tRNA identity." Nucleic Acids Res **26**(22): 5017-35.

Goodman, H. M., J. Abelson, et al. (1968). "Amber suppression: a nucleotide change in the anticodon of a tyrosine transfer RNA." Nature **217**(5133): 1019-24.

Heim, R. and R. Y. Tsien (1996). "Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer." Curr Biol **6**(2): 178-82.

Hohsaka, T. and M. Sisido (2002). "Incorporation of non-natural amino acids into proteins." Curr Opin Chem Biol **6**(6): 809-15.

Ibba, M. and H. Hennecke (1995). "Relaxing the substrate specificity of an aminoacyl-tRNA synthetase allows in vitro and in vivo synthesis of proteins containing unnatural amino acids." FEBS Lett **364**(3): 272-5.

Jevsevar, S., M. Kunstelj, et al. "PEGylation of therapeutic proteins." Biotechnol J **5**(1): 113-28.

Johnson, J. A., Y. Y. Lu, et al. "Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications." Curr Opin Chem Biol **14**(6): 774-80.

Kobayashi, T., O. Nureki, et al. (2003). "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion." Nat Struct Biol **10**(6): 425-32.

Kobayashi, T., T. Takimura, et al. (2005). "Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase." J Mol Biol **346**(1): 105-17.

Kurschus, F. C., P. P. Pal, et al. (2009). "Gold fluorescent annexin A5 as a novel apoptosis detection tool." Cytometry A **75**(7): 626-33.

Lepthien, S., M. G. Hoesl, et al. (2008). "Azatryptophans endow proteins with intrinsic blue fluorescence." Proc Natl Acad Sci U S A **105**(42): 16095-100.

Link, A. J., M. L. Mock, et al. (2003). "Non-canonical amino acids in protein engineering." Current Opinion in Biotechnology **14**(6): 603-609.

Liu, D. R., T. J. Magliery, et al. (1997). "Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo." Proc Natl Acad Sci U S A **94**(19): 10092-7.

Liu, W., A. Brock, et al. (2007). "Genetic incorporation of unnatural amino acids into proteins in mammalian cells." Nat Methods **4**(3): 239-44.

Lund, E. and J. E. Dahlberg (1998). "Proofreading and aminoacylation of tRNAs before export from the nucleus." Science **282**(5396): 2082-5.

Merkel, L., M. G. Hoesl, et al. "Blue fluorescent amino acids as in vivo building blocks for proteins." Chembiochem **11**(3): 305-14.

Myslinski, E., J. C. Ame, et al. (2001). "An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene." Nucleic Acids Res **29**(12): 2502-9.

Neihardt, F. C., J. Parker, et al. (1975). "Function and regulation of aminoacyl-tRNA synthetases in prokaryotic and eukaryotic cells." Annu Rev Microbiol **29**: 215-50.

Neumann, H., A. L. Slusarczyk, et al. "De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs." J Am Chem Soc **132**(7): 2142-4.

Noren, C. J., S. J. Anthony-Cahill, et al. (1989). "A general method for site-specific incorporation of unnatural amino acids into proteins." Science **244**(4901): 182-8.

Nowak, M. W., J. P. Gallivan, et al. (1998). "In vivo incorporation of unnatural amino acids into ion channels in Xenopus oocyte expression system." Methods Enzymol **293**: 504-29.

Quinn, C. L., N. Tao, et al. (1995). "Species-specific microhelix aminoacylation by a eukaryotic pathogen tRNA synthetase dependent on a single base pair." Biochemistry **34**(39): 12489-95.

Sakamoto, K., A. Hayashi, et al. (2002). "Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells." Nucleic Acids Res **30**(21): 4692-9.

Schimmel, P. (1987). "Aminoacyl tRNA synthetases: general scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs." Annu Rev Biochem **56**: 125-58.

Schimmel, P., R. Giege, et al. (1993). "An operational RNA code for amino acids and possible relationship to genetic code." Proc Natl Acad Sci U S A **90**(19): 8763-8.

Schimmel, P. and C. C. Wang (1999). "Getting tRNA synthetases into the nucleus." Trends Biochem Sci **24**(4): 127-8.

Sharp, S., D. DeFranco, et al. (1981). "Internal control regions for transcription of eukaryotic tRNA genes." Proc Natl Acad Sci U S A **78**(11): 6657-61.

Shoulders, M. D., K. A. Satyshur, et al. "Stereoelectronic and steric effects in side chains preorganize a protein main chain." Proc Natl Acad Sci U S A **107**(2): 559-64.

Steer, B. A. and P. Schimmel (1999). "Major anticodon-binding region missing from an archaebacterial tRNA synthetase." J Biol Chem **274**(50): 35601-6.

Tcherkezian, J., P. A. Brittis, et al. "Transmembrane receptor DCC associates with protein synthesis machinery and regulates translation." Cell **141**(4): 632-44.

Thibodeaux, G. N., X. A. Liang, et al. (2010). "Transforming a Pair of Orthogonal tRNA-aminoacyl-tRNA Synthetase from Archaea to Function in Mammalian Cells." Plos One **5**(6): -.

Umeda, A., G. N. Thibodeaux, et al. (2009). "Site-specific Protein Cross-Linking with Genetically Incorporated 3,4-Dihydroxy-L-Phenylalanine." Chembiochem **10**(8): 1302-1304.

Wakasugi, K., C. L. Quinn, et al. (1998). "Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant." EMBO J **17**(1): 297-305.

Wang, L., A. Brock, et al. (2001). "Expanding the genetic code of Escherichia coli." Science **292**(5516): 498-500.

Wang, L., T. J. Magliery, et al. (2000). "A New Functional Suppressor tRNA/Aminoacylâˆ’tRNA Synthetase Pair for the in Vivo Incorporation of

Unnatural Amino Acids into Proteins." Journal of the American Chemical Society **122**(20): 5010-5011.

Wang, L. and P. G. Schultz (2005). "Expanding the genetic code." Angewandte Chemie-International Edition **44**(1): 34-66.

Wang, L., Z. Zhang, et al. (2003). "Addition of the keto functional group to the genetic code of Escherichia coli." Proc Natl Acad Sci U S A **100**(1): 56-61.

Wang, W. Y., J. K. Takimoto, et al. (2007). "Genetically encoding unnatural amino acids for cellular and neuronal studies." Nature Neuroscience **10**(8): 1063-1072.

Weinmann, R., H. J. Raskas, et al. (1974). "Role of DNA-dependent RNA polymerases II and III in transcription of the adenovirus genome late in productive infection." Proc Natl Acad Sci U S A **71**(9): 3426-39.

Wold, F. (1981). "In vivo chemical modification of proteins (post-translational modification)." Annu Rev Biochem **50**: 783-814.

Xie, J., L. Wang, et al. (2004). "The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination." Nat Biotechnol **22**(10): 1297-301.

Yaremchuk, A., I. Kriklivyi, et al. (2002). "Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition." EMBO J **21**(14): 3829-40.

Ye, S., C. Kohrer, et al. (2008). "Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis." J Biol Chem **283**(3): 1525-33.

Yoo, T. H., A. J. Link, et al. (2007). "Evolution of a fluorinated green fluorescent protein." Proc Natl Acad Sci U S A **104**(35): 13887-90.

Zhang, Z., L. Alfonta, et al. (2004). "Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells." Proc Natl Acad Sci U S A **101**(24): 8882-7.

Zhang, Z., B. A. Smith, et al. (2003). "A new strategy for the site-specific modification of proteins in vivo." Biochemistry **42**(22): 6735-46.

Zhu, J., C. Lu, et al. (2008). "Single mutation on the surface of Staphylococcus aureus Sortase A can disrupt its dimerization." Biochemistry **47**(6): 1667-74.

Vita

Liang Xiang was born in Beijing, China to Dr. T. Xiang and Dr. Y. Gao. He attended James E. Taylor High School in Katy, Texas before enrolling in the Electrical Engineering program at the University of Illinois in Urbana-Champaign, Illinois. After his freshmen year he transferred to Northwestern University in Evanston, Illinois where he graduated with a B.S. in Biomedical Engineering in 2007. In 2008 he rolled in the Biomedical Engineering PhD program at the University of Texas at Austin under the supervision of Dr. Zhiwen Jonathan Zhang.

As a summer intern for Royal Dutch Shell he has worked on the design and programming of a pressure-volume-temperature analysis program. Some other past projects include designing rehabilitation devices for patients after a stroke,
a fetal heart rate monitor for beluga whales, and EZ Plasmid Mapper (http://infosake.com/plasmid/index.html). He is primarily interested in the application of biomedical engineering to solve modern medical problems.

Permanent email: lxiang21@gmail.com
This dissertation was typed by Liang Xiang